Research Repository UCD

Title	Recommendations for Biomarker Identification and Qualification in Clinical Proteomics
Authors(s)	Mischak, H., Allmaier, G., Apweiler, R., et al.
Publication date	2010-08-25
Publication information	Mischak, H., G. Allmaier, R. Apweiler, and et al. “Recommendations for Biomarker Identification and Qualification in Clinical Proteomics.” American Association for the Advancement of Science, August 25, 2010. https://doi.org/10.1126/scitranslmed.3001249.
Publisher	American Association for the Advancement of Science
Item record/more information	http://hdl.handle.net/10197/5014
Publisher's version (DOI)	10.1126/scitranslmed.3001249

Downloaded 2026-05-01 23:36:08

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Recommendations for Biomarker Identification and Qualification in Clinical Proteomics

Harald Mischak,*† Günter Allmaier, Rolf Apweiler, Teresa Attwood, Marc Baumann, Ariela Benigni, Samuel E. Bennett, Rainer Bischoff, Erik Bongcam-Rudloff, Giovambattista Capasso, Joshua J. Coon, Patrick D'Haese, Anna F. Dominiczak, Mohammed Dakna, Hassan Dihazi, Jochen H. Ehrich, Patricia Fernandez-Llama, Danilo Fliser, Jorgen Frokiaer, Jerome Garin, Mark Girolami, William S. Hancock, Marion Haubitz, Denis Hochstrasser, Rury R. Holman, John P.A. Ioannidis, Joachim Jankowski, Bruce A. Julian, Jon B. Klein, Walter Kolch, Theo Luider, Ziad Massy, William B. Mattes, Franck Molina, Bernard Monsarrat, Jan Novak, Karlheinz Peter, Peter Rossing, Marta Sánchez-Carbayo, Joost P. Schanstra, O. John Semmes, Goce Spasovski, Dan Theodorescu, Visith Thongboonkerd, Raymond Vanholder, Timothy D. Veenstra, Eva Weissinger, Tadashi Yamamoto, Antonia Vlahou

*Author affiliations can be found in the supporting online material.
†Corresponding author. E-mail: mischak@mosaiques-diagnostics.com

Recommendations about structuring proteomic biomarker studies should increase the probability that such markers will be clinically useful.

ABSTRACT

Clinical proteomics has yielded some early positive results—the identification of potential disease biomarkers—indicating the promise for this analytical approach to improve the current state of the art in clinical practice. However, the inability to verify some candidate molecules in subsequent studies has led to skepticism among many clinicians and regulatory bodies, and it has become evident that commonly encountered shortcomings in fundamental aspects of experimental design mainly during biomarker discovery must be addressed in order to provide robust data. In this Perspective, we assert that successful studies generally use suitable statistical approaches for biomarker definition and confirm results in independent test sets; in addition, we describe a brief set of practical and feasible recommendations that we have developed for investigators to properly identify and qualify proteomic biomarkers, which could also be used as reporting requirements. Such recommendations should help put proteomic biomarker discovery on the solid ground needed for turning the old promise into a new reality.

INTRODUCTION

One of the aims of proteomics—the analysis of the complete set of proteins in a given specimen (e.g. biological fluid or tissue)—is to identify biomarkers of disease: molecules that are consistently modified or present at abnormal concentrations in specific illnesses or other health conditions, as graphically depicted in Fig. 1. Markers might be

important for determining susceptibility to disease or foretelling disease progression, or might predict the effect of a particular treatment on clinical outcomes. Such clinical proteomics initially raised high hopes after reports were published of potentially informative biomarkers for several diseases. Some of these biomarkers have moved to the stage of testing in larger populations (*1-3*) and in clinical trials (*4*). However, many early claims of candidate molecules were not substantiated in subsequent studies (*5-7*). This situation has created skepticism about the value of clinical proteomics. We argue that the failures of these initial studies reflect shortcomings in fundamental aspects of the experimental designs (*8, 9*). Our Perspective seeks to offer guidance to investigators on how to structure future studies so as to increase the likelihood that proteomic biomarkers will be clinically useful.

In the last few years, efforts have been made to offer preliminary guidance for clinical proteomic studies (*10, 11*). Some recommendations have already been adopted by key journals (*12, 13*). Unfortunately, even simple, common-sense standards are not widely followed, and today's literature on newly discovered biomarkers still contains a substantial number of potentially invalid or even misleading reports. Reasons for spurious findings include (i) small sample sizes; (ii) small effects; (iii) lack of standardization in designs, definitions, outcomes, and analytical modes; and (iv) fragmentation of efforts through multiple teams of investigators (*14*). Most proteomic biomarker studies in the past have succumbed to these flaws, but these can be avoided in future studies.

The strategies and tools required to identify dozens or even hundreds of biomarkers from a pool of thousands of features are not yet common knowledge among biomarker investigators. As a consequence, during their second annual meeting (*15*) members of the European Kidney and Urine Proteomics program (EuroKUP) initiated an effort to create minimal, generally applicable, concise guidelines for designing and executing discovery studies of clinical proteomic biomarkers and reporting their results. Several additional scientists active in the field of clinical proteomics and methodologists have joined this effort. The target audience for the recommendations in this Perspective includes scientists, reviewers, editors, clinicians, and funding bodies. Most previous guidelines on diverse research study designs and topics have focused on the proper reporting of research (*16, 17*), but reporting is just a mirror of the design and conduct of a study. It is important to apply appropriate methods and to provide sufficient details about these methods so that other researchers may evaluate the results and reproduce them; the latter is critically important for biomarker qualification and further development.

DEFINITIONS

To facilitate implementation of these recommendations, we have defined the major terms and kept the recommendations straightforward and broadly applicable, regardless of the analytical platform, type of biological samples investigated, or epidemiological design (case-control, cohort, cross-sectional, nested study, or other). For reporting on general aspects of epidemiological study design and diagnostic performance, researchers can consult the STROBE (*18*) and STARD (*19*) statements, whenever applicable. We focus here in more detail on aspects that are specific or peculiar to clinical proteomic research.

We define a "proteomic biomarker" as a specific peptide or protein that is associated with a specific condition, such as the onset, manifestation, or progression of a disease or a response to treatment. Similarly, a "biomarker profile" is defined as a combination of distinct proteomic biomarkers, using a clearly defined algorithm giving a readout parameter associated with the specific condition. We denote a biomarker as "qualified" on the basis of evidence for the above-mentioned association, generally dependent on appropriate statistics (see below). The concept of qualifying biomarkers for purposes such as drug development has been discussed (*20*). In the current Perspective, we emphasize developing the appropriate evidence that links the biomarker with biological processes and/or clinical endpoints. It is important to match a biomarker's purpose with the evidence required for qualifying the biomarker for that purpose (*21*). For example, a biomarker for the prognosis of diabetic nephropathy must be prospectively examined in a population not harboring clinically evident disease at the time of sampling, and its value must be assessed according to its ability to predict future clinical outcome. Claiming potential prognostic value simply on the basis of examination of the biomarkers in patients with disease in comparison with nondiseased individuals is inappropriate (*22*).

"Biomarker qualification" is a conclusion that within the stated context of use, the results of assessment with a biomarker can be relied on to have a specific interpretation and application in drug development and regulatory decision-making. The context of use of a biomarker is a comprehensive statement that fully and clearly describes the manner and purpose of use for the biomarker. The context-of-use statement describes all important criteria regarding the circumstances under which the biomarker is qualified.

A biomarker is defined by its "associative" aspect. It has no causal or mechanistic relevance beyond that, unless additional functional/biological evidence supports these claims (*23*). In diseases with inherently broad clinical spectra (such as cancer), a biomarker may be linked to only a particular variant or stage of the disease or to specific complications such as metastasis (*24*). Qualified biomarkers will be valuable tools to diagnose patients earlier in the clinical course of disease, to predict outcomes and guide interventional approaches, to stratify participants in clinical trials, and to monitor response to therapy, and may be used in personalized medicine. Thus, it is imperative to clarify up front and report the appropriate clinical setting in which the association applies.

Here, we focus on the initial identification of a biomarker and the minimal requirements for qualification and reporting biomarker discovery findings. Initial promising reports do not suffice for turning a biomarker or biomarker profile into a widely used clinical test. This process requires validation and proper evaluation of the test performance (sensitivity, specificity, and positive and negative predictive values) in specific settings and demonstration of clinical utility, applicability, and cost-effectiveness.

MAJOR ISSUES IN DESIGN AND REPORTING

The major issues that need to be addressed in the design and reporting of clinical biomarker studies include the following:

Clearly defined clinical question, outcomes, and selection of subjects. The precise clinical question, the pertinent outcomes, and the purpose (potential clinical

application) of the biomarker must be clearly stated. Appropriate positive and negative controls must be defined, as deemed proper for the clinical question. Healthy controls are often inappropriate for defining disease-specific biomarkers; controls with related or similar diseases must be examined. For instance, for the discovery of biomarkers for diabetic kidney disease the appropriate control cohort would be age-matched diabetic patients without kidney disease.

Clinical conditions and outcomes are not always assessed with certainty. Standardized, widely acceptable criteria for the documentation of outcomes should be used. The intended spectrum of disease and potential of outcome misclassification should be reported. Samples from individuals with unclear diagnoses may be omitted in the discovery and early qualification stage, at which most studies assume a case-control design (the comparison of otherwise similar individuals with and without a given condition). However, in real-life clinical practice, some people may unavoidably belong in this gray zone, potentially affecting the performance of a qualified biomarker.

The assessment of patients whose diagnosis is based on surrogate parameters (such as isolated microalbuminuria—a small increase in albumin in the urine, which can be an early indication of diabetic kidney disease—as surrogate for diabetic renal damage) should be avoided; clinically accepted events and hard endpoints should ideally be used.

Sometimes the presence of a particular disease is not determined by using the same criteria in all participants. For example, prostate cancer is documented by examining the prostate after its removal, whereas healthy controls cannot have the prostate removed to exclude occult cancer-disease that is clinically unsuspected, which is a frequent occurrence for this type of cancer. Such verification bias is difficult to correct (*25, 26*) and should be clearly acknowledged.

Sufficient demographic and/or clinical data. A study attempting to define biomarkers must be accompanied by appropriate clinical information (required to properly confirm clinical status) and demographic and phenotypic data about the subjects from whom the specimens have been collected. These data should at least include age, gender, ethnic background, and detailed status of the disease or condition under investigation as well as relevant physiological parameters (such as blood pressure or body mass index), comorbidities, and current medications or treatment.

Sufficient information about the sampling methodology. A detailed description of specimen collection, including how and when the samples were obtained, handled, and stored, with a detailed description of containers and stabilizing solutions (such as anticoagulants or protease inhibitors) must be provided. These parameters should be appropriate and applied consistently. If not, the quality of samples and consequently the resulting data are compromised. When practical considerations do not allow optimal sampling methodology, this fact should be clearly documented and the limitations acknowledged.

Sufficient information about the experimental methodology. Mass spectrometry (MS)–based techniques for identifying proteins in clinical samples usually encompass a separation step, using techniques such as liquid chromatography or electrophoresis to

resolve individual proteins, before MS analysis. The resolution of the separation and the resolution and mass accuracy of the MS analysis must be proportional to the complexity of the investigated sample, which is high for most biomarker-discovery studies. Poor MS resolution and accuracy or the absence of feature annotation, which will preclude identification of the claimed biomarkers, is not acceptable.

For some metabolic diseases, a well-described single biomarker is accurate for the diagnosis or monitoring the clinical course. However, in situations of complex pathophysiology (such as diabetes-associated vascular disease) a single biomarker may not provide optimal diagnostic accuracy. In these situations, a panel of biomarkers appears to be more appropriate (*10, 27*). A biomarker profile must consist of clearly defined molecular entities. Ideally, the precise chemical composition of the biomarkers should be known. In the case of proteins and peptides, this information includes the complete amino acid sequence of the potential biomarker (which, because of proteolytic modification, is generally not identical to the corresponding entry in databases, such as SwissProt) and all posttranslational modifications (PTMs). Presently, this goal cannot always be achieved, owing to limitations in technology. Thus, in current practice biomarkers are defined mainly by physical parameters, such as molecular mass, migration/retention time in separation, or interaction with an antibody. Amino acid sequences are usually derived from database entries, which often miss essential information about frequently observed PTMs. The exact chemical entity has been unambiguously identified in only a few cases (*28*). In light of these constraints, a sequence from the database in the absence of the exact chemical definition or physicochemical properties in the absence of sequence should be considered acceptable for now. However, bearing in mind that both definitions are incomplete, the analytical platform must permit assessment of the defined chemical entity (and its relative amount) with satisfactory confidence.

The platform characteristics and performance of the entire analytical procedure (sample collection, sample preparation, etc.) should be known and adequately described in order to assess the quality of any data sets. To determine the analytical variability of the analytical platform, a set of experiments should be performed to determine features including intra-assay and inter-assay precision, temperature stability of analytes, postpreparation stability, and time course of sampling as outlined in detail in guidance documents (*29*). If these experiments have already been done in previous studies, provision of appropriate references to them suffices. If these parameters are not properly determined, the level of confidence in the analysis and, consequently, the relevance of the biomarker cannot be assessed.

To attribute the same identity to a certain feature in several independent analyses, accepted deviations of mass and other parameters (retention time, migration, position on gel, etc.) must be reported. When repeatedly analyzing the same sample, the observed deviation in identifying parameters and (relative) abundance (the precision) should be reported. In addition, steps should be included to ensure that a biomarker measurement is not compromised by the occurrence of disease-associated changes in PTMs. If PTMs cannot be addressed by the applied analytical technology, additional steps—such as PTM-

specific blotting or lectin blotting (to detect specific sugar moieties)—should be undertaken.

Appropriate statistical approaches, including adequate sample size, proper adjustments, and correction for multiple testing. The statistical analysis should account for technical variability, biological variability, confounding factors, and other anticipated sources of bias: which aspects of the analyses were predefined, and which arose as post hoc exploratory analyses should be clarified.

The selection of sample size should be justified on the basis of rational calculations of statistical power. If sample size is based only on constraints of sample availability, this should be acknowledged. As a result of high biological variation, a sufficient number of independent samples is essential to estimate the correct distribution of a potential biomarker. The investigation of <12 samples does not allow even a reasonable estimation of the mean and variance based on the normal distribution (*30*) and should be avoided. Small, underpowered studies are subject to an enhanced risk of both false negatives (not finding a true association) and false positives (finding a spurious association). When samples are unavoidably limited in single centers (such as in cases of rare and orphan diseases), the appropriate approach is to initiate a multi-center collaborative study. Finally, pooling of samples compromises statistical assessment and eliminates information on outliers and variability. There can be high between-subject variance in proteomics datasets, hence pooling should generally be avoided.

Statistical analyses should evaluate at least model calibration and discrimination performance, and they also should evaluate reclassification performance whenever pertinent. Calibration refers to the goodness of fit across the range of predicted risks; discrimination assesses how well those with an outcome are separated from those without the outcome of interest. Reclassification examines whether a new prognostic model places participants in more appropriate risk categories as compared with an older model. Details of methods should be provided for each of these purposes.

Adjustments for potential covariates—variables that might be correlated with the investigated pathophysiology—should be justified and clarified whether they are considered a priori or post hoc, and if so, why. As an example, age influences the risk of coronary artery disease. Thus, data sets should either include individuals of similar age in cases and controls, or if this is not possible, adjustment for age-related alterations in the distribution of potential biomarkers must be made. If patients can be treated for the condition under study, this capacity should be taken into account in the design and analysis, keeping in mind whether the biomarker is intended for use in treated patients, untreated patients, or both. Both biomarkers and outcomes can be influenced by therapeutic measures (for example, drugs). Evaluating untreated individuals (for instance, samples from the placebo group of a clinical trial) may yield results different from those for treated individuals.

Numerous reports have demonstrated the importance of strict and correct use of statistics, especially the need to adjust for multiple testing (*31, 32*) to reduce false associations. For example, the simultaneous testing of 1000 potential biomarkers at a level

of $P = 0.05$ will yield the erroneous identification of approximately 50 spurious biomarkers by chance. Adjustments for multiple testing correct for this fact (*33*). There are many approaches that can properly account for multiplicity, including not only frequentist methods, but also false-discovery and Bayesian approaches. The selected methods should be properly documented and appropriate references provided. If adjustment for multiple testing is not made when investigating multidimensional data sets, the findings are probably false and generally not relevant. If significant disease-associated changes cannot be found when applying appropriate statistics, the solution cannot be avoidance of correction for multiple testing (leaving sound scientific ground). Instead, either more samples should be analyzed in order to obtain better estimates of the relevance of the investigated biomarkers, or it must be accepted that significant biomarkers cannot be identified in the context under study.

Various machine-learning algorithms allow the combination of multidimensional data sets and perform similarly (and very well) in establishing multi-marker models (*34, 35*). However, these algorithms lead to meaningful results only if the number of data sets is sufficient to uncover the latent structure in the data and allow for generalization. The parameters used in the machine-learning algorithms should be clearly stated so that analyses can be reproduced by other scientists.

Confirmation in independent test sets. The observation of a significant association in a given data set does not ensure that the findings can be generalized in other data sets or that the association is highly specific for the investigated condition. Most statistical approaches used for biomarker evaluation assume (i) an even distribution of features across the data (similar variance in control and disease groups, and the absence of covariates), (ii) that the findings can be generalized, and (iii) that an association exists only with the investigated condition. These simplifications are generally not correct. As a consequence, most biomarkers with promising results in a first data set will turn out to have less promising results in independent data sets, as demonstrated recently (*3, 36*). Validation failure indicates that additional unidentified latent variables are, in part, responsible for the observed differences between cases and controls and/or that the selected features were false positive or nongeneralizable. Generally, the strength of associations is expected to be inflated in the original, discovery analysis (*37*). Consequently, qualification of any biomarker or panel in at least one independent test set (samples that were not used for the initial identification of the biomarker) is essential to avoid reporting flawed conclusions. This concern applies to both single biomarkers and biomarker profiles but is even more important for the latter because the algorithms generally over-fit the data, resulting in a model that relies on information/potential biomarkers that were coincidentally found to be altered in the data set but that may not have any true correlation with disease. Over-fitting can result in nearly 100% sensitivity and specificity in a training set (the data set used to initially identify biomarkers and establish an apparently disease-associated model), even upon cross-validation, but classification accuracy often decreases sharply in an independent test set, particularly when the size of the training set has been small (*38*).

To maximize generalizability, external validation should be performed, whenever feasible, preferably in samples from multiple sites, such as in a multi-center setting. Ideally,

evaluation of validation samples should be performed in blinded fashion. The sampling of subjects and the characteristics of the validation sample (or samples) need to be described in the same detail as for those of the discovery data set. Outcomes, experimental methodology, and statistical analysis should be the same in the test and validation samples; any unavoidable deviations should be acknowledged and discussed in order to clarify the effect on the results.

RECOMMENDATIONS

We recommend that the guidelines listed in Table 1 become standard requirements for the scientific reporting of proteomic biomarker data. Adherence to these requirements does not imply that biomarkers emanating from an analysis can automatically be adopted for clinical application or that a valid test has been described. However, if the association of a defined biomarker with a specific pathophysiological condition has been properly evaluated, even on an analytical platform that is probably not applicable for a clinical setting (such as two-dimensional gel electrophoresis followed by MS), then such results are worthy of publication, regardless of whether the results are "negative" or "positive." Communication of "positive" results should allow their reproduction and further qualification by other teams; eventually, one should hopefully be able to develop assays for clinical applications on the basis of well-qualified biomarkers. Brief communication of "negative" results is also necessary to avoid loss of resources and time by additional scientists who would otherwise follow the same inefficient paths of investigation and to avoid the distortion of the literature by publication bias. If no significance of the proposed biomarker can be demonstrated, then that should be clearly stated. Terms like "potential biomarkers" should be avoided because they do not contain any important information.

CONCLUSION

This Perspective offers guidance to investigators on how to structure and report future studies to increase the likelihood that proteomic biomarkers will be clinically useful. The scientific community should accept findings about such markers only if the requirements related to biological specimen collection, data reporting, and identification of biomarkers are met. Scientific journals may use these recommendations to assess manuscripts submitted for publication. It is particularly important that biomarker investigations be performed in compliance with these recommendations for the recent Biobanking and Biomolecular Resources Research Infrastructure initiatives (*39*) that aim to secure sustainable access to biological resources required for health-related research and development. These samples represent valuable resources that cannot be replenished and thus must not be wasted. Although this paper presents the view of more than 50 scientists in the field, the recommendations are open for additional discussion. For this reason, we have opened a portal (*40*) in which individuals are encouraged to post their comments and present their opinions.

Clinical proteomics, if properly applied, may enable major progress in clinical medicine from which many patients would ultimately benefit. However, to reach this goal standards for quality and scientific validity in clinical proteomics studies and their reports must be ensured.

Fig. 1. Clinical proteomics workflow. Biological specimens, mostly body fluids such as urine or blood (but also tissue), are collected, and their proteome is examined in great detail in order to identify proteins and peptides that are significantly altered in disease. Upon validation of the results in independent cohorts, these can be considered as biomarkers, enabling improvement in, for example, diagnosis, prognosis, and management of patients. In addition, these biomarkers—via their association with pathophysiology—hold the promise to guide development of improved therapeutic drugs.

CREDIT: C. BICKEL/*SCIENCE TRANSLATIONAL MEDICINE*

Table 1. Requirements for scientific reporting of proteomic biomarker data.

Describe and justify the clinical question, outcomes, and selection of subjects	Describe the clinical question and justify why it is of interest; describe what outcomes are assessed and comment on their clinical validity, potential for misclassification and verification bias, if pertinent; clarify what are the eligibility criteria for the selected study populations and justify specific choices.
Describe the assessed subjects	Provide demographic information with gender, age, ethnic origin, and concomitant medications at a minimum, and all relevant disease-related and clinical parameters.
Describe sampling	Provide an accurate description of the sampling conditions and procedures (including the collection process and any manipulation of the sample before storage, the time between sampling and storage, storage conditions, and the addition of any protease inhibitors and/or preservatives). Justify the sampling choices according to the literature or supporting experimental data.
Describe experimental methodology	The procedure, as well as the observed standard deviation of technical specifications related to the procedure, should be given. To attribute the same identity to a certain feature in several independent analyses, accepted deviations of mass and other parameters (retention time, migration, position on gel, etc.) must be reported. Also, the observed deviation in identifying parameters and (relative) abundance, when the same sample is analyzed repeatedly, must be reported.
Describe the statistical evaluation	Provide details on determination of sample size, statistical analysis plan (for appraising calibration, discrimination, and/or reclassification), any consideration or adjustment for covariates (including treatment, whenever pertinent), methods

	for adjustment for multiplicity, and parameters used in complex machine-learning approaches, whenever pertinent. Clarify which analyses are predefined and which are post hoc.
Validate results	The results must be confirmed in at least one independent sample set. The sampling and characteristics of the validation population should be reported, and the analysis should be symmetrical in the test and validation data sets; any deviations should be reported.
Acknowledge limitations	No study is perfect; limitations should be clearly acknowledged and their potential impact on the results discussed.
Take responsibility	The contributions of each author should be clearly stated.

References and Notes

1. S. Decramer, S. Wittke, H. Mischak, P. Zürbig, M. Walden, F. Bouissou, J. L. Bascands, J. P. Schanstra, Predicting the clinical outcome of congenital unilateral ureteropelvic junction obstruction in newborn by urinary proteome analysis. *Nat. Med.* **12**, 398–400 (2006).

2. E. M. Weissinger, E. Schiffer, B. Hertenstein, J. L. Ferrara, E. Holler, M. Stadler, H. J. Kolb, A. Zander, P. Zürbig, M. Kellmann, A. Ganser, Proteomic patterns predict acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Blood* **109**, 5511–5519 (2007).

3. K. Rossing, H. Mischak, M. Dakna, P. Zürbig, J. Novak, B. A. Julian, D. M. Good, J. J. Coon, L. Tarnow, P. Rossing; PREDICTIONS Network, Urinary proteomics in diabetes and CKD. *J. Am. Soc. Nephrol.* **19**, 1283–1290 (2008).

4. http://www.controlled-trials.com/ISRCTN03911524/ISRCTN03911524.

5. N. Rifai, M. A. Gillette, S. A. Carr, Protein biomarker discovery and validation: The long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983 (2006).

6. A. Alaiya, M. Al-Mohanna, S. Linder, Clinical cancer proteomics: Promises and pitfalls. *J. Proteome Res.* **4**, 1213–1222 (2005).

7. E. F. Petricoin, III, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, L. A. Liotta, Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577 (2002).

8. D. McLerran, W. E. Grizzle, Z. Feng, W. L. Bigbee, L. L. Banez, L. H. Cazares, D. W. Chan, J. Diaz, E. Izbicka, J. Kagan, D. E. Malehorn, G. Malik, D. Oelschlager, A. Partin, T. Randolph, N. Rosenzweig, S. Srivastava, S. Srivastava, I. M. Thompson, M. Thornquist, D. Troyer, Y. Yasui, Z. Zhang, L. Zhu, O. J. Semmes, Analytical

validation of serum proteomic profiling for diagnosis of prostate cancer: Sources of sample bias. *Clin. Chem.* **54**, 44–52 (2008).

9. J. Hu, K. R. Coombes, J. S. Morris, K. A. Baggerly, The importance of experimental design in proteomic mass spectrometry experiments: Some cautionary tales. *Brief. Funct. Genomics Proteomics* **3**, 322–331 (2005).

10. H. Mischak, R. Apweiler, R. E. Banks, M. Conaway, J. J. Coon, A. Dominizak, J. H. Ehrich, D. Fliser, M. Girolami, H. Hermjakob, D. F. Hochstrasser, V. Jankowski, B. A. Julian, W. Kolch, Z. Massy, C. Neususs, J. Novak, K. Peter, K. Rossing, J. P. Schanstra, O. J. Semmes, D. Theodorescu, V. Thongboonkerd, E. M. Weissinger, J. E. Van Eyk, T. Yamamoto, Clinical proteomics: A need to define the field and to begin to set adequate standards. *Proteomics Clin. Appl.* **1**, 148–156 (2007).

11. M. A. Knepper, Common sense approaches to urinary biomarker study design. *J. Am. Soc. Nephrol.* **20**, 1175–1178 (2009).

12.http://www3.interscience.wiley.com/homepages/112770559/2456_instruc.pdf.

13. http://www.mcponline.org/cpmeeting/cguidelines.pdf.

14. J. P. Ioannidis, Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).

15. A. Vlahou, G. Allmaier, T. Attwood, E. Bongcam-Rudloff, A. Charonis, J. Frokiaer, H. Mischak, J. Schanstra, G. Spasovski, 2nd combined working group and management committee meeting of urine and kidney proteomics COST action. *Proteomics Clin. Appl.* **3**, 1017–1022 (2009).

16. A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, M. Vingron, Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).

17. D. Moher, S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, D. G. Altman, CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, (mar23 1), c869 (2010).

18. E. von Elm, D. G. Altman, M. Egger, S. J. Pocock, P. C. Gøtzsche, J. P. Vandenbroucke; STROBE Initiative, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).

19. P. M. Bossuyt, J. B. Reitsma, D. E. Bruns, C. A. Gatsonis, P. P. Glasziou, L. M. Irwig, J. G. Lijmer, D. Moher, D. Rennie, H. C. de Vet; Standards for Reporting of Diagnostic Accuracy, Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* **326**, 41–44 (2003).

20. J. A. Wagner, Strategic approach to fit-for-purpose biomarkers in drug development. *Annu. Rev. Pharmacol. Toxicol.* **48**, 631–651 (2008).

21. C. A. Altar, D. Amakye, D. Bounos, J. Bloom, G. Clack, R. Dean, V. Devanarayan, D. Fu, S. Furlong, L. Hinman, C. Girman, C. Lathia, L. Lesko, S. Madani, J. Mayne, J. Meyer, D. Raunig, P. Sager, S. A. Williams, P. Wong, K. Zerba, A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clin. Pharmacol. Ther.* **83**, 368–371 (2008).

22. H. Mischak, P. Rossing, Proteomic biomarkers in diabetic nephropathy—Reality or future promise? *Nephrol. Dial. Transplant.* 22 June 2010 (10.1093/ndt/gfq363).

23. F. Goodsaid, J. E. Bandow, H. Mischak, Grand rounds in proteomics at the FDA White Oak, Silver Spring, MD, USA, April 3, 2007. *Proteomics Clin. Appl.* **1**, 1526–1531 (2007).

24. J. W. Lee, V. Devanarayan, Y. C. Barrett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J. A. Rogers, R. Millham, P. J. O'Brien, J. Sailstad, M. Khan, C. Ray, J. A. Wagner, Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* **23**, 312–328 (2006).

25. M. Buzoianu, J. B. Kadane, Adjusting for verification bias in diagnostic test evaluation: A Bayesian approach. *Stat. Med.* **27**, 2453–2473 (2008).

26. J. H. Page, A. Rotnitzky, Estimation of the disease-specific diagnostic marker distribution under verification bias. *Comput. Stat. Data Anal.* **53**, 707–717 (2009).

27. S. Decramer, A. Gonzalez de Peredo, B. Breuil, H. Mischak, B. Monsarrat, J.-L. Bascands, J. P. Schanstra, Urine in clinical proteomics. *Mol. Cell. Proteomics* **7**, 1850–1862 (2008).

28. D. M. Good, V. Thongboonkerd, J. Novak, J. L. Bascands, J. P. Schanstra, J. J. Coon, A. Dominiczak, H. Mischak, Body fluid proteomics for biomarker discovery: Lessons from the past hold the key to success in the future. *J. Proteome Res.* **6**, 4549–4555 (2007).

29. www.labcompliance.com/info/links/methods/guidelines.aspx.

30. R. V. Hogg, E. A. Tanis, *Probability and Statistical Interference* (McMillian, New York, 2010).

31. S. Dudoit, M. J. van der Laan, *Multiple Testing Procedures with Applications to Genomics* (Springer, New York, 2008).

32. D. I. Broadhurst, D. B. Kell, Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171–196 (2006).

33. H. Mischak, J. J. Coon, J. Novak, E. M. Weissinger, J. P. Schanstra, A. F. Dominiczak, Capillary electrophoresis-mass spectrometry as a powerful tool in biomarker discovery and clinical diagnosis: An update of recent developments. *Mass Spectrom. Rev.* **28**, 703–724 (2009).

34. A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, S. Levy, A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. *Bioinformatics* **21**, 631–643 (2005).

35. S. Dudoit, J. Fridlyand, T. P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* **97**, 77–87 (2002).

36. M. Haubitz, D. M. Good, A. Woywodt, H. Haller, H. Rupprecht, D. Theodorescu, M. Dakna, J. J. Coon, H. Mischak, Identification and validation of urinary biomarkers for differential diagnosis and dvaluation of therapeutic intervention in ANCA associated vasculitis. *Mol. Cell. Proteomics* **8**, 2296–2307 (2009).

37. J. P. Ioannidis, Why most discovered true associations are inflated. *Epidemiology* **19**, 640–648 (2008).

38. U. M. Braga-Neto, E. R. Dougherty, Is cross-validation valid for small-sample microarray classification? *Bioinformatics* **20**, 374–380 (2004).

39. http://www.bbmri.eu/.

40. http://www.eurokup.org/node/132.

41. Funding: This initiative was supported in part by the EuroKUP COST-Action (BM0702; www.eurokup.org), by the European Community's 6th Framework Programme, grant agreement LSHM-CT-2006-037093 (InGenious HyperCare), and by the European Community's 7th Framework Programme, grant agreement HEALTH-F2-2009-241544 (SysKID). A.F.D. wishes to acknowledge support of the British Heart Foundation Chair and Programme Grants (BHF RG/07/005/23633). B.A.J. and J.N. acknowledge their support in part from NIH grants DK075868, DK078244, DK082753, DK083663, and DK080301. B.M. and J.P.S. acknowledge the support from the Agence Nationale pour la Recherche (ANR-07-PHYSIO-004-01), the Fondation pour la Recherche Médicale "Grands Equipements pour la Recherche Biomédicale," and the CPER2007-2013 program. J.J. was supported by a grant from Federal Ministry of Education and Research (01GR0807). W.H. acknowledges support by NCI grant CA 128427 and Korean WCU grant R31-2008-000-10086-0. A.V. and J.G. acknowledge support from FP7 DECanBio (grant agreement 201333). OJS acknowledges support from NIH/NCI CA CA085067
Competing interests: H.M. is the co-founder and co-owner of Mosaiques-Diagnostics.

Citation: H. Mischak, G. Allmaier, R. Apweiler, T. Attwood, M. Baumann, A. Benigni, S. E. Bennett, R. Bischoff, E. Bongcam-Rudloff, G. Capasso, J. J. Coon, P. D'Haese, A. F. Dominiczak, M. Dakna, H. Dihazi, J. H. Ehrich, P. Fernandez-Llama, D. Fliser, J. Frokiaer, J. Garin, M. Girolami, W. S. Hancock, M. Haubitz, D. Hochstrasser, R. Holman, J. P. A. Ioannidis, J. Jankowski, B. A. Julian, J. B. Klein, W. Kolch, T. Luider, Z. Massy, W. B. Mattes, F. Molina, B. Monsarrat, J. Novak, K. Peter, P. Rossing, M. Sánchez-Carbayo, J. P. Schanstra, O. J. Semmes, G. Spasovski, D. Theodorescu, V. Thongboonkerd, R. Vanholder,

T. Veenstra, E. Weissinger, T. Yamamoto, A. Vlahou, Recommendations for biomarker identification and qualification in clinical proteomics. *Sci. Transl. Med.* **2**, 46ps42 (2010).

Supporting Online Material

www.sciencetranslationalmedicine.org/cgi/content/full/2/46/46ps42/DC1

Author affiliations