DRUG DISCOVERY TODAY BIOSILICO

Practical proteomic biomarker discovery: taking a step back to leap forward

Jennifer Listgarten and Andrew Emili

There is a pressing need for radically improved proteomic screening methods that allow for earlier diagnosis of disease, for systematic monitoring of physiological responses and for uncovering the fundamental mechanisms of drug action. Recent developments in proteomic technology offer tremendous, yet untapped, potential to yield novel biomarkers that are translatable to routine clinical use. Despite the significant conceptual promise of comparative proteomic profiling as a research platform for biomarker discovery, however, major hurdles remain for practical and clinical implementation. In particular, there is growing recognition that rigorous experimental design principles are urgently required to validate conclusively the unproven methodologies currently being touted. Debate and confusion persist about where the burden of proof lies: statistically, biologically or clinically? Moreover, there is no consensus about what constitutes a meaningful benchmark. An important question is how to achieve a scientifically rigorous, and therefore convincing, proof-of-concept that can be accepted by the field. Key analytical challenges related to these issues that must be addressed by the burgeoning biomarker community are discussed here.

Jennifer Listgarten
Department of Computer Science,
University of Toronto,
Toronto, Ontario,
M5S 3G4, Canada
Andrew Emili*
Program in Proteomics and Bioinformatics,
Banting and Best Department of Medical Research,
University of Toronto,
Toronto, Ontario,
M5G 1L6, Canada
*e-mail: andrew.emili@utoronto.ca

Although early detection of pathology can greatly improve patient outcomes, there are relatively few good noninvasive molecular tools for diagnosis of early-stage disease, such as premalignant cancer [1,2], or for prognosis, such as identifying cardiac patients at higher risk of dying (e.g. the Framingham Heart Study), or of non-responsiveness to therapy [3]. The development of satisfactory therapeutics is also hampered by the lack of suitable bioassays that evaluate drug efficacy or toxicity [4]. Therefore, there is currently a pressing need to develop methodology that allows for routine and reliable identification and stringent validation of molecular indicators – biomarkers – in readily accessible patient samples, such as blood, urine or sputum. Such biomarkers could provide a method for clinical monitoring of broad classes of common illnesses (e.g. cancer, heart disease and infection), for assessing the effectiveness of therapeutic intervention regimes and for improving the reliability and accuracy of clinical trials. However, because of considerable disease heterogeneity, inter-patient variation and other irrelevant sources of biological variability, biomarker discovery is extremely challenging. Moreover, to a certain extent, biomarker discovery presupposes knowledge of which variables, such as irrelevant epiphenomena, account for confounding factors.

Because biochemical responses to disease or drug action are likely to be reflected in the patterns of protein expression and turnover in affected cells,

1359-6446/05/$ – see front matter PII: S1359-6446(05)03645-7

tissues and, presumably, blood, proteomic profiling using either targeted [5] or unbiased global [6] protein identification and quantification is expected to provide insight into the pathophysiological changes that precede or accompany clinical presentation. Although numerous promising immunoassay-based proteomic methods have been developed in recent years (e.g. miniaturized and multiplexed readout systems), protein mass spectrometry (MS) has emerged, because of its versatility, sensitivity and accuracy, as a technology of choice for identifying potentially clinically useful molecular patterns in cancer [2], heart disease [7] and other common ailments. The pioneering cancer profiling studies by Petricoin *et al.* [8,9] suggested that the detection of proteomic 'signatures' in serum by mass-spectrometry-based screening could offer a substantial improvement over existing diagnostic strategies in terms of sensitivity and specificity. However, despite widespread initial optimism [10–12], the actual clinical impact of large-scale high-throughput proteomic technologies on biomedical investigations of disease and drug action has been limited to date, and substantial questions have been raised regarding the validity of proposed screening procedures [3,13–19].

One could argue that the main reason for the present failure to produce a convincing biomarker discovery experiment stems not from technological limitations in detecting trace amounts of disease-specific molecules, but rather from fundamental difficulties in elucidating whether putative biomarker patterns are truly clinically informative. In other words, conclusively determining whether systematic proteomic measurements of state-specific fluctuations in the levels of protein components in biological fluids can be translated into clinical practice in an effective, consistent and meaningful manner. However, as Ransohoff [20] so rightly points out, 'we cannot decide whether new things do work by reasoning about whether they should work' – that is to say, the debate should be strictly evidence-based, even if guided by scientific intuition.

Despite recent progress, a plethora of important unresolved analytical issues continue to hamper progress in the field [10,21,22]. These difficulties are often compounded by the fact that a limited number of primary tissue samples is usually available for pilot screening studies, and the multivariate proteomic datasets generated are usually extremely complex and subject to nonstatic biological influences. Nevertheless, despite documented and perceived limitations in current proteomic screening methodologies, the prospect of systematic biomarker discovery could be within reach. Given careful consideration and resolution of key outstanding challenges, this rapidly emerging field could be well poised for spinning proteomic patterns into biomarker gold.

Evoking the evidence

Fundamentally, the goal of biomarker discovery is to find a distinctive molecular signal with a clear-cut clinical value. For example, earlier detection of disease [23], improved patient stratification and better monitoring of therapeutic intervention [3], are all key milestones towards the establishment of personalized medicine [24]. Blood-based protein assays are widespread in clinical practice and generally have a meritorious, if somewhat spotty, record in terms of patient impact. Some of the more prominent success stories include the detection of liberated cardiac muscle troponin for diagnosis of myocardial infarction [25], the monitoring of heart failure patients based on circulating brain natriuretic peptide levels [26] and the screening of prostate-specific antigen levels in serum from males at risk of prostate cancer [27].

To date, most clinically useful blood markers have been found either serendipitously or through careful evaluation of individual candidate proteins based on hypotheses regarding a particular disease. As with many other aspects of biomedical research, current efforts in biomarker discovery have been mostly centered on high throughput MS-based profiling of human-derived samples, such as blood and tissue biopsies [23]. Yet, despite the pioneering and attention-grabbing reports of the preliminary findings of putative ovarian and prostate cancer signatures in blood using simple MS-based protein screening of patient serum combined with sophisticated statistical and machine learning algorithms [8,9], successful translation of proteomic screening technology into clinical and pharmaceutical practice has been lacking to date.

The lack of a definitive demonstration of practical utility, combined with a growing awareness of possible pitfalls in current experimental protocols, has led to critical voicing of the imperative for cogent benchmarks to substantiate platforms and claims [3,13–17]. Skeptics now abound because of the lack of a clear consensus about what constitutes rigorous experimental design and validation. A major impediment to developing a convincing biomarker discovery pipeline is the concern that proteomic approaches are predisposed to uncharacterized bystander effects; that is, the detection of irrelevant epiphenomenon stemming from extraneous experimental factors, such as spurious differences in the patient cohort like diet or environment.

What the biomarker research community urgently needs is an airtight, carefully designed and well-executed experiment that conclusively demonstrates proof-of-concept. The combined assortment of proteomic platforms, computational tools and biological samples currently available might be sufficient for effective biomarker discovery, but the field needs to take a collective pause to define more clearly how experiments can effectively and convincingly target the questions on hand and to address current criticisms [3,13–19], thereby convincing the skeptics. Such a step will be the bridge to faster progress toward the application of biomarker profiling in a real clinical setting.

Patients are out of control

We endorse the view of Hartwell and colleagues [28], who have proposed the use of well-defined experimentally tractable model organisms, such as inbred laboratory mouse strains, to validate proteomic profiling procedures for the early detection of cancer – or other disease – in blood samples. Animal models offer a tremendous advantage because disease-affected tissue can be generated and accessed under tightly controlled experimental conditions that minimize genetic or environmental differences, which confound biomarker discovery in a heterogeneous outbred human population. Although the candidate biomarkers identified in a transgenic or knockout mouse might not be orthologous to those in human patients, the potential for rigorously establishing proof-of-principle of the discovery process in such a well-controlled experimental setting is tremendous. Because the entire sample history, from genetic background through to phenotyping and ending in collection of target tissue and blood, can be more completely documented and accounted for, a well-designed experiment in this domain would be far more credible than the current standard of practice commonly found in human trials.

If successful, such a landmark profiling study would serve to validate the technical procedures used to generate the proteomic datasets and the entire data analysis pipeline. This would set the stage for embarking on the greater challenge of genuine biomarker discovery using human clinical specimens, which are far more likely to be subject to factors beyond methodical investigative control.

Biomarker catch-22

The prevalent use of healthy human samples as a control group in biomarker discovery studies is an obvious, but perhaps unwise, choice. Predictive models (e.g. a classifier such as a support vector machine) usually pick out a dominant signal over a more subtle signal. These terms (dominant/subtle) are defined, crucially, with respect to the case and control groups together. If the dominant signal distinguishing between ovarian cancer patients and healthy volunteers is, for example, a general inflammatory signature (such as a systemic acute phase response) then a subtler ovarian-cancer-specific signal might never be uncovered or appropriately modeled by such a comparative profiling experiment.

Inevitably, the selection of appropriate controls could be a costly, time-consuming, iterative process. Suppose, for example, that in a first-pass serum screening experiment of breast cancer samples together with closely matched (to the extent that this is known at the time of experimentation) healthy case controls, only a systemic acute-phase signal (or some other nonspecific response) is detected. Then one knows that this factor needs to be taken into consideration, and a new set of controls, which is thought to be better matched in this respect, must be used in the next phase of the study; for example, a different cohort of cancer patients also thought to present with an acute-phase response. Note that such a protocol requires identification and biological understanding of putative preliminary biomarkers. It is not entirely clear how one can avoid a seemingly endless experimental protocol.

An alternative, but probably unfeasible, approach that does not require expert biological knowledge of the putative biomarkers, would be to use an expansive control group representative of the broader population that might regularly undergo the type of diagnostic test being considered. If sufficiently representative, such a population would be expected to include healthy individuals, as well as individuals at various stages of all common ailments. Of course, gathering and screening such a comprehensive set of specimens is not practical, especially for smaller, preliminary pilot studies.

Although randomized controlled trials (the gold standard of clinical drug trials) might seem at first glance to be a natural remedy to this difficult situation, the search for early-detection biomarkers is not amenable to such a setup, because one cannot dole out the seeds of cancer to one group in the same way that one can assign different drug regimens to two groups of patients (although this could potentially be done in animals). That is in contrast to drug efficacy clinical trials, in which cause can be controlled to affect change; biomarker discovery is inherently an observational science. Thus, the search for biomarker discovery lies in very difficult terrain. However, the effort to outline and apply the kind of rigorous experimental and statistical guidelines that are now established in clinical drug trials [29] to the area of biomarker discover [30] is a very valuable approach. Nonetheless, much work remains in drawing up analogous standardized biomarker discovery guidelines partly, as mentioned above, because of the different nature of this fledgling area.

What you seek is what you get

If what ones seeks is an early marker for, for example, some particular cancer of interest, then one must obtain samples from patients either long before their cancer is diagnosed (but who then go on to develop the cancer of interest) or, at the very least, who are in the early stages of disease. Other samples might not work as a reasonable proxy because the biological hallmarks of cancer are in most cases likely to progress with stage. Although late stage pathology might contain a (possibly embedded) signal related to the early onset of cancer, this signal could be overshadowed by more dominant late-stage expression patterns. However, many biomarker studies use an *ad hoc* mixture of stages, typically with very few early stage cancers. These are, by definition, harder to come by in the absence of reliable diagnostics and therefore such profiling might never elucidate a signal of interest, if it exists. It is those diseases currently lacking an effective early detection screening method that are also most likely to benefit from a successful biomarker project, yet these same diseases are

the ones for which it is hardest to obtain early stage samples (in fact, this can only be done prospectively). Animal models offer a potential solution.

Burden of proof is in the pudding, not the toppings

What kind of statistical or machine learning methodologies can we use or develop to guarantee that we have found the biomarker we are seeking? Consider the following situation: you are provided with broad access to unlimited samples taken from two specific classes of your choice – a cancer and a control group. You run thousands of samples from each group through a mass spectrometer, and then hand the resulting data to your favorite machine learning specialist or biostatistician, who applies sophisticated preprocessing followed by state-of-the-art predictive modeling. They return and tell you that, based on the best statistical and machine learning methodology (and suppose, for the sake of argument, that they know the methodology perfectly, and that the methodology is perfect) they estimate that any future samples you give to them, drawn from either of the same two populations and processed in an identical way, could be accurately predicted with a sensitivity and specificity of 100%.

A 'eureka' moment for biomarker discovery, perhaps? But what does such an experiment actually validate? Regardless of what algorithms were used, theoretically, what kind of meaningful clinical information can be garnered from such a study? Can the entire biomarker pipeline be validated? No. We can conclude that there is some informative signal – disease-specific perhaps, but possibly not – that can be detected accurately, and then used to reliably predict new samples if drawn from one of the same two populations. The difficult, but crucial, questions about whether this signal is disease specific, generalizable to the screening population at large (unless this happened to be exactly the control population) and of any biological consequence, can never be answered by machine learning or statistics alone. Statistical and machine learning methods are not a crutch for poor experimental design nor, crucially, can they elucidate fundamental insight from poorly designed experiments.

Overfitting is not an inherent limitation

A common staple of the biomarker literature is a statement to the effect that overfitting has occurred, or that overfitting might have occurred, or that overfitting is always a problem with large-scale high throughput data-driven experimental paradigms, particularly with regards to biomarker detection. Although it is certainly true that improper use of machine learning and statistical techniques can lead to overfitting, it is also true that this should not be so and that this problem is easily avoided. Claiming that overfitting is an inherent limitation in the biomarker context is akin to stating that sloppy laboratory work is an inherent limitation in biology; although sloppy bench habits do exist, these are easily identified and overcome. A technical discussion of how to avoid overfitting is beyond the scope of this review; however, we briefly point out that there are two main approaches: the use of cross-validation (or resampling-based methods) and the use of Bayesian methodology[1] [31,32].

Machine learning buzzwords aren't going to make or break the field of biomarker discovery

Articles on biomarker discovery whose sole novel contribution is presenting an existing fancy-sounding predictive machine learning algorithm X that has not yet been presented to the biomarker community offer little intellectual contribution, and do little to advance the state-of-the-art in biomarker discovery. It is true that different algorithms will find different features and perhaps perform differently [33] but the bottleneck, at present, does not result from inability to obtain good prediction performance. Rather, it is the entire experimental design and validation process that is in dire need of fresh new ideas before machine learning and statistics can truly play an interesting role.

This is made evident by the fact that reported predictive accuracies are extremely high or even perfect [13,34–37]; we do not need to improve upon these accuracies, but rather to question whether these perfect results were generated using suitable test data [18,19,21,33,38]. Of course, when one is sure that the data is appropriate, one can then consider optimizing the choice of predictive algorithms to achieve the necessarily extremely high sensitivity and specificity that allow for reasonable positive predictive value in the clinic. Additionally, many of the methods currently in use, for example, linear support vector machines, logistic regression and neural networks without a hidden layer, are small variations of one another and thus publication of a putative new method in the context of a proteomic dataset should be justified or at least discussed in this context.

Instrumentation: focus on the calibration

It is now clear that the surface-enhanced laser desorption and ionization time-of-flight (SELDI-TOF) MS instrumentation used in the pioneering studies by Liotta and colleagues [8,9] had insufficient resolution to allow for the unambiguous identification of the putative marker molecules, which is needed if they are to be validated (ideally) using an alternate methodology or for forming the basis of a simplified, more widely adopted diagnostic [16]. Although much attention has been paid to the need for meticulous

[1]Cross-validation and resampling-based methods are essentially based on extensions to the idea of a 'hold out' dataset. Bayesian methodology intuitively limits model complexity using probabilistic parameter averaging within a model class so that a highly complex model cannot pick out a single best parameter setting – as can happen in many non-Bayesian paradigms – if there is not enough data to warrant it. Essentially, Bayesian methodology automatically adapts the complexity of a model so that it is not greater than is warranted by the data (and in so doing, avoids overfitting).

sample preparation [39,40], relatively scant attention has been given to the need for vigilance in calibrating existing analytical systems so that they are appropriately tailored for clinical pattern recognition. The newest generations of mass spectrometer instruments (particularly the so-called hybrids) are very sensitive, highly accurate and fast scanning, but they remain subject to artifacts imposed by signal-to-noise restrictions, skewed precision and systematic biases that lead to misinterpretation and under-sampling [21,33] – limitations that are often underappreciated by non-experts. Measurement uncertainty is further confounded by the extreme complexity of most biological fluids, in particular blood with its daunting dynamic range window [10,40].

Given the current emphasis in the field on sample concentration, fractionation and depletion, insufficient attention is given to testing platform reproducibility and practical detection limits in a pertinent manner [18], such as documenting the linearity or reproducibility of responses of a tunable range of spiked protein standard in a germane target mixture. Likewise, although the rapid advances in hardware development made in the past few years have resulted in perceived performance attributes dominating much of the current scientific discussion, a solid understanding of their limitations has yet to emerge, with few qualified estimates of how much of the proverbial biomarker iceberg currently remains below the proteomic surface.

Benchmarking the benchmarks

At this point, the field would benefit from a collective target or benchmark to stimulate future development. But what constitutes a suitable benchmark for this type of problem? One might consider a call for a generic experiment to assess new prediction algorithms like the CASP (critical assessment of techniques for protein structure prediction) or CAPRI (critical assessment of predicted interactions) exercises do for protein structure prediction [41,42]. However, the problem of biomarker discovery is somewhat different. Whereas protein structure prediction has a clearly defined input and output (given, for example, X-ray crystallography as a gold standard) with only a computational algorithm in between, biomarker discovery encompasses a diverse (albeit integrated) series of steps. Each step is dependent on the other, starting with a decision about which samples are relevant for collection and processing, followed by high-throughput measurement and only then being tackled with computational and statistical methods.

It can even be argued that any benchmark data other than raw patient samples can not serve as a true benchmark because it has been subjected to part of the methodological protocol decisions. Additionally, there is no gold standard for biomarker discovery – spike-in experiments serve at best as a consolation 'bronze' standard. As such, it is unlikely that there can be a single feasible and conclusive type of benchmarking exercise. Rather, benchmarking for biomarker discovery will be more loosely defined, most likely constituting a series of experiments, each more convincing than the next, or separately elucidating the validity of different components, until at long last, a putative biomarker pipeline is shown to work in a realistic clinical setting.

However, more intermediate, although still quite comprehensive benchmark-type experiments could certainly be helpful. For example, in a recent inspiring experiment, Semmes *et al.* [43] reported on inter-laboratory calibration of SELDI platforms, followed by largely successful prediction of cancer versus control at the various laboratory sites when using the same raw samples.

There is an overarching need for a suite of benchmarking exercises, ranging from those such as the Semmes *et al.* [43] study, to calibration style benchmarking where, for example, the linearity of instrument responsiveness is established [44], to the ultimate benchmark – real clinical usage – as well as for many challenges in between, such as data normalization, peak detection, identification and quantification and, at some point, classification.

Onwards we march

The biomarker field, currently a disparate, rapidly evolving and somewhat confused research domain, can draw strength and motivation from successes in large-scale clinical trials (but should also heed recent cautionary tales). Although the critical definition of 'standards of excellence' in biomarker research remains open to debate, sound perspectives regarding appropriate computational procedures, data interpretations and reporting mechanisms have recently been put forward [45]. We have pointed out a basic difference between the problems of determining drug efficacy and safety and that of discovering a biomarker, although the two problems are similar in that, ultimately, the only way to truly validate a biomarker will be with large and expensive clinical trials that directly establish true biomedical value. However, thoughtfully designed intermediate experiments will hasten us toward such successes. Despite the ongoing controversy over the merits of high throughput-based biomarker discovery, and regardless of the depth of the challenges before us, the biomarker community can steer to where the true path lies by stepping back every now and then to critically reassess implied and explicit default assumptions.

References

1 Etzioni, R. *et al.* (2003) The case for early detection. *Nat. Rev. Cancer* 3, 243–252
2 Hanash, S. (2004) Integrated global profiling of cancer. *Nat. Rev. Cancer* 4, 638–644
3 Stoughton, R.B. and Friend, S.H. (2005) How molecular profiling could revolutionize drug discovery. *Nat. Rev. Drug Discov.* 4, 345–350
4 Frank, R. and Hargreaves, R. (2003) Clinical biomarkers in drug discovery and development. *Nat. Rev. Drug Discov.* 2, 566–580
5 Anderson, L. (2005) Candidate-based proteomics in the search for biomarkers of cardiovascular disease. *J. Physiol.* 563, 23–60
6 Hanash, S. (2003) Disease proteomics. *Nature* 422, 226–232
7 Stanley, B.A. *et al.* (2004) Heart disease, clinical proteomics and mass spectrometry. *Dis. Markers* 20, 167–178
8 Petricoin, E.F. 3rd *et al.* (2002) Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst.* 94, 1576–1578

9 Petricoin, E.F. *et al.* (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577
10 Veenstra, T.D. *et al.* (2005) Biomarkers: mining the biofluid proteome. *Mol. Cell. Proteomics* 4, 409–418
11 Conrads, T.P. *et al.* (2003) Cancer diagnosis using proteomic patterns. *Expert Rev. Mol. Diagn.* 3, 411–420
12 Liotta, L.A. *et al.* (2003) Clinical proteomics: written in blood. *Nature* 425, 905
13 Sorace, J.M. and Zhan, M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics* 4, 24
14 Diamandis, E.P. (2004) Proteomic patterns to identify ovarian cancer: 3 years on. *Expert Rev. Mol. Diagn.* 4, 575–577
15 Diamandis, E.P. (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer Inst.* 96, 353–356
16 Diamandis, E.P. (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol. Cell. Proteomics* 3, 367–378
17 Diamandis, E.P. (2003) Re: Serum proteomic patterns for detection of prostate cancer. *J. Natl. Cancer Inst.* 95, 489–90; author reply 490–491
18 Baggerly, K.A. *et al.* (2005) Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J. Natl. Cancer Inst.* 97, 307–309
19 Hu, J. *et al.* (2005) The importance of experimental design in proteomic mass spectrometry experiments: some cautionary tales. *Brief Funct. Genomic Proteomic* 3, 322–331
20 Ransohoff, D.F. (2004) Evaluating discovery-based research: when biologic reasoning cannot work. *Gastroenterology* 127, 1028
21 Nesvizhskii, A.I. and Aebersold, R. Interpretation of shotgun proteomics data: The protein inference problem. *Mol. Cell. Proteomics* (in press)
22 Pan, S. *et al.* (2005) High throughput proteome screening for biomarker detection. *Mol. Cell. Proteomics* 4, 182–190
23 Johann, D.J., Jr *et al.* (2004) Clinical proteomics and biomarker discovery. *Ann. N. Y. Acad. Sci.* 1022, 295–305
24 Jain, K.K. (2002) Personalized medicine. *Curr. Opin. Mol. Ther.* 4, 548–558
25 Hamm, C.W. *et al.* (1997) Emergency room triage of patients with acute chest pain by means of rapid testing for cardiac troponin T or troponin I. *N. Engl. J. Med.* 337, 1648–1653
26 Weber, T. *et al.* (2005) The diagnostic and prognostic value of brain natriuretic peptide and aminoterminal (nt)-pro brain natriuretic peptide. *Curr. Pharm. Des.* 11, 511–525
27 Catalona, W.J. *et al.* (1991) Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* 324, 1156–1161
28 Aebersold, R. *et al.* (2005) Perspective: a program to improve protein biomarker discovery for cancer *J. Proteome Res.* 4, 1104–1109
29 *International Conference on Harmonisation of Technical Requirements for Regulation of Pharmaceuticals for Human Use* ICH Harmonized Tripartite Guideline: Statistical principles for clinical trials (1999)
30 Manne, U. *et al.* (2005) Recent advances in biomarkers for cancer diagnosis and treatment. *Drug Discov. Today* 10, 965–976
31 Gelman, A. *et al.* (2004) *Bayesian Data Analysis* (2nd edn), CRC Press
32 Hastie, T. *et al.* (2001) *The Elements of Statistical Learning*, Springer
33 Listgarten, J. and Emili, A. (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 4, 419–434
34 Baggerly, K.A. *et al.* (2003) A comprehensive approach to the analysis of matrix-assisted laser desorption/ionization-time of flight proteomics spectra from serum samples. *Proteomics* 3, 1667–1672
35 Petricoin, E.F. *et al.* (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572–577
36 Qu, Y. *et al.* (2002) Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin. Chem.* 48, 1835–1843
37 Satten, G.A. *et al.* (2004) Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. *Bioinformatics* 20, 3128–3136
38 Conrads, T.P. and Veenstra, T.D. (2005) What have we learned from proteomic studies of serum? *Expert Rev Proteomics* 2, 279–281
39 Kislinger, T. *et al.* (2005) Multidimensional Protein Identification Technology (MudPIT): Technical Overview of a Profiling Method Optimized for the Comprehensive Proteomic Investigation of Normal and Diseased Heart Tissue. *J. Am. Soc. Mass Spectrom.* 16, 1207–1220
40 Bjorhall, K. *et al.* (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. *Proteomics* 5, 307–317
41 Moult, J. (2005) A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* 15, 285–289
42 Wodak, S.J. and Mendez, R. (2004) Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* 14, 242–249
43 Semmes, O.J. *et al.* (2005) Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin. Chem.* 51, 102–112
44 Wang, W. *et al.* (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* 75, 4818–4826
45 Lyons-Weiler, J. (2005) Standards of excellence and open questions in cancer biomarker research: an informatics perspective. *Cancer Informatics* 1, 1–7