Clinical Chemistry 62:4
551–557 (2016)

Q&A

Proteogenomics: Opportunities and Caveats

Moderators: Lampros Dimitrakopoulos,[1] Ioannis Prassas,[2] and Eleftherios P. Diamandis[1,2,3*]
Experts: Alexey Nesvizhskii,[4] Thomas Kislinger,[5] Jacob Jaffe,[6] and Andrei Drabovich[7,8]

Proteogenomics is a rapidly evolving field at the intersection of genomics, transcriptomics, and proteomics. Whole genome, exome, and RNA sequencing are well-established techniques that can provide information at the DNA and RNA level with excellent sequencing coverage and depth. Although tens of thousands of clinical samples have been sequenced thus far, data integration and interpretation still remain largely incomplete. Recent advances in proteomic technologies have enabled the accurate and almost complete characterization of the proteomes of many tissues and biological fluids. Integration of multiomics data for the accurate annotation and reciprocal refinement of genomic and proteomic models is essentially the goal of proteogenomics.

This integrative approach has the potential to provide solid evidence for the translation of previously unknown transcripts. Those transcripts and the respective encoded proteins might be implicated in physiological or pathophysiological processes. Novel reported peptides can represent single amino acid variants, splice variants, gene fusions, RNA editing events, novel open reading frames, translated noncoding RNAs, and pseudogenes, among many others. Proteogenomic platforms can now be used to investigate which of these novel "events" gets translated at the protein level, thereby implicating them as candidate new druggable targets or as new diagnostic or prognostic biomarkers for a wide spectrum of diseases.

The potential for such identifications is maximized when both sequencing and raw proteomic data originate from the very same sample under investigation. It is becoming clear that this "sample-specific" approach, and the use of matched customized search databases, is associated with lower false-positive and false-negative identification rates. However, like all areas of active research, proteogenomics in its current state is not free of drawbacks. Major limitations in the field are the sensitivity of the mass spectrometers, the increased false discovery rate for the novel peptide hits, and the inherent biophysical properties that render some peptides undetectable.

In this Q&A we discuss with 4 experts in the field the current status of proteogenomics and conditions that have to be met to deliver its promises.

What are the key technologies that enabled the development of proteogenomics?

Alexey Nesvizhskii: In most cases, especially when studying human or model organisms, proteogenomics is critically dependent on the knowledge (and often aims to refine that knowledge) assembled by large genome and proteome annotation teams that build genome-centric resources such as Ensemble and RefSeq and protein-level resources such as the UniProt knowledge base (UniProtKB).[9] Thus, proteogenomics is critically dependent on those efforts and the technologies that they use. With respect to experiment-specific data used as part of proteogenomics studies, on the genomics side it commonly involves next generation sequencing (NGS) data such as exome se-

[1] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; [2] Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; [3] Department of Clinical Biochemistry, University Health Network, Toronto, Ontario, Canada; [4] Associate Professor of Pathology, Associate Professor of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; [5] Associate Professor, Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; [6] Director, LINCS Proteomic Characterization Center for Signaling and Epigenetics, Associate Director, Proteomics Platform, The Broad Institute, Cambridge, MA; [7] Department of Clinical Biochemistry, University Health Network, Toronto, Ontario, Canada; [8] Assistant Professor, Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada.

* Address correspondence to this author at: 60 Murray St., 6th Floor, Toronto, On, Canada M5G 1X5. Fax 416-586-8628; e-mail ediamandis@mtsinai.on.ca.
Received October 19, 2015; accepted December 7, 2015.

[9] Nonstandard abbreviations: UniProtKB, UniProt knowledge base; NGS, next generation sequencing; MS/MS, tandem mass spectrometry; WGS, whole genome sequencing; UTR, untranslated region; uORF, upstream open reading frames; SNV, single-nucleotide variant; SNP, single-nucleotide polymorphisms; MRM-MS, multiple reaction monitoring–mass spectrometry; TCGA, The Cancer Genome Atlas; CPTAC, Clinical Proteomic Tumor Analysis Consortium; SRM, selected reaction monitoring; FDR, false discovery rate; ICGC, International Cancer Genome Consortium.

quencing and transcriptomics (RNA-Seq). Because proteogenomics is most dependent on the availability of high-quality proteomic data, the technologies that enable sensitive, large-scale proteome profiling are of the highest importance. Tandem mass spectrometry (MS/MS) is the most dominant technology for high-throughput quantitative proteome profiling. The most commonly used strategy is to digest proteins into peptides using an enzyme such as trypsin, followed by MS/MS sequencing of the resulting peptides. This variant of proteomics is called shotgun (or bottom-up) proteomics. Intact proteins can also be analyzed using mass spectrometry (top-down proteomics), and such data can naturally be used as part of proteogenomics studies as well, but for technical reasons top-down proteomics has yet to enter the proteomics mainstream. Ribosomal proofing is a promising technology that provides complementary information at the level of translational products. This technology is at present technically challenging for most laboratories. However, such data are extremely valuable for proteogenomics applications as they allow more direct linkage between transcriptomics and proteomics data. Last but not least, bioinformatics is of critical importance. This includes databases and public repositories for storing raw data, tools for processing the data coming from each technology, and tools for integration and visualization of such data. In particular, the development of proteomics data repositories such as PeptideAtlas and ProteomeXchange was critical for proteogenomics because they provided computational scientists interested in proteogenomics access to large proteomics data sets needed for their work.

Thomas Kislinger: Proteogenomics was first described over a decade ago and has played a crucial role in the development and refinement of genome models for a variety of organisms. Renewed interest in this type of an approach is certainly a result of the rapid improvements in sequencing technologies [NGS, exome-Seq, whole genome sequencing (WGS), RNA-Seq], but likely more closely tied to technological innovations allowing for the characterization of deep proteome profiles from small biological/clinical samples. These more comprehensive data sets have renewed hope that, alongside appropriate bioinformatics and statistical frameworks, novel biologically relevant results can be gleaned from proteogenomic data.

Jacob Jaffe: The now routine nature of genome sequencing is an important enabler for proteogenomics today. But even when genome sequencing was relatively more difficult, the data could be leveraged for early proteogenomic applications. Today, proteogenomics and gene expression profiling are becoming even more aligned with the development of RNA-Seq and ribosomal footprinting technologies, enabling better mapping to gene products in specific cellular contexts.

Andrei Drabovich: Genome-wide next-generation DNA and RNA sequencing, high-resolution mass spectrometry, sample-specific customized protein databases, bioinformatic algorithms, and software tools to integrate multiple omics data sets enable proteogenomics and facilitate its use for practical applications. As an example, high-resolution mass spectrometry with hybrid quadrupole-time-of-flight or quadrupole-Orbitrap instruments allows for high-throughput analysis of many thousand proteins and provides accurate peptide sequencing data, thus reducing the number of false-positive matches. Such methods eventually help in the identification of rare peptide variants, such as cancer-specific missense mutations.

What is the additional information that proteogenomics can offer compared to either genomic or proteomic platforms?

Alexey Nesvizhskii: Using a narrower definition of the term proteogenomics, proteogenomics studies aim to validate or refine gene models and annotations produced using genomic data. Existing databases such as Ensemble already contain many annotated transcripts whose protein coding potential is unknown, i.e., sequences predicted based on their genome annotation pipelines but with little or no previous experimental evidence of their expression at the protein level. Similarly, the UniProtKB database—the most commonly used reference database for proteomics studies, contains sequences divided into categories (evidence levels), including the most dubious proteins that have been assigned to P4 and P5 categories.

By matching tandem mass spectra against the sequences of those proteins, one can identify their presence in biological samples. Thus, proteomics-based evidence in the form of identified peptides, and in combination with other relevant data that may be available (sequence conservation, RNA-Seq, ribosomal profiling, etc.), can be used to improve the annotation of the corresponding transcripts/proteins. As the most desirable scenario, proteogenomics analysis would result in a promotion of a previously questionable transcript to the status of experimentally confirmed protein-coding sequence. Similarly, proteogenomics can provide protein-level evidence for a particular variant or isoform, e.g., an amino acid variant or an alternative splice form.

In a broader sense, proteogenomics refers to all sorts of applications involving joint analysis of sample-specific genomics/transcriptomics and proteomics data. Studies in which NGS data, such as exome and especially RNA-Seq transcriptomic data, and proteomics data generated in parallel are becoming increasingly common. In such studies, sample-specific genomic and transcriptomic data can be used to reconstruct the transcriptomes of the samples under investigation. Such transcriptomes are inherently noisy, i.e., they contain sequence reads suggesting thousands of novel or uncharacterized events (novel splice junctions, sequence variants, chimeric transcripts, noncoding RNAs, gene fusions, RNA editing events, etc.). Using genomics and transcriptomics data, one can then build a custom protein sequence database containing, in addition to known sequences taken from a reference sequence database, many predicted, novel sequences. By matching mass spectrometry data against this custom database one can obtain sample specific, protein-level evidence of expression for a subset of those events, likely bringing the biological or clinical significance of those events to a higher level.

Thomas Kislinger: In combination, the coding proteome and noncoding transcriptome represent the end products of the sequence-to-phenotype continuum (DNA to RNA to Protein). The emerging view is that proteomic and transcriptomic approaches provide complementary readouts of the cellular state with neither holding a monopoly over defining the molecular phenotype. Of course, proteomics comes with the technical caveat that current technologies are not sensitive enough to identify every expressed protein sequence (at least the problem is more pronounced than in transcriptomics). Therefore, by combining genomic, transcriptomic and proteomic technologies, in a proteogenomic workflow, these technologies can inform each other. Classically, proteogenomics provides definitive protein-centric proof for the expression of a given DNA or RNA sequence (and perhaps more importantly, the mutant variants of these sequences). This peptide centric evidence can help refine current gene models and improve current reference protein sequence databases. As proteogenomics evolves beyond simply validating genomic predictions about the proteome, we will surely discover that the parameters perturbed to generate these predictions in the first-place can be reoptimized using evidence based on proteomic detection.

Jacob Jaffe: Proteogenomics is both a complement to genomics and new paradigm for interpretation and visualization of proteomics data. In an increasingly genomics-centered world, proteomics (as a field) does itself a service by putting its data onto a scaffold that genomics folks can easily understand. Meanwhile, the underlying proteomic data can reveal things about biology that are inaccessible to genomics. How powerful is it to see that a phosphorylation site is recurrently mutated in certain cancers? That's the power of proteogenomics.

Andrei Drabovich: Proteogenomics facilitates confirmation and correction of existing genes or even identification of potentially new genes. Current standard proteomic platforms rely on the reference genomic sequences and thus miss polymorphisms, mutated proteins, and rare protein variants. With multiple mechanisms leading to such rare variants, I would highlight peptides expressed by pseudogenes and noncoding RNAs as the most exciting area in proteogenomics, with the potential to identify some rare and even novel biological mechanisms. Proteogenomic data may also offer more reliable prioritization of cancer driver genes compared to genomic platforms, as was recently demonstrated for colon cancer.

What type of variant peptides can be detected by proteogenomics (which are currently missed by classical proteomics)?

Alexey Nesvizhskii: There is very long list of novel peptides that can potentially be identified using proteogenomics. These include novel splice junctions, peptides containing single amino acids variants, and peptides corresponding to alternative start sites. Other rare events include RNA editing events and gene fusions. In principle, one can detect peptides mapping to intergenic regions suggesting novel open reading frames, or to regions currently annotated as pseudogenes and noncoding RNA. Another category of novel peptides is peptides mapping to known protein-coding regions, including to their untranslated regions (UTRs), but in an alternative frame [e.g., peptides derived from upstream open reading frames (uORFs)]. Many recent studies reported identifications of all sorts of novel peptides mentioned above. Unfortunately, most of those studies did not apply the level of stringency in filtering their data that is required

for detection of low-likelihood events (in most cases, the same filtering criteria were applied to the detection of known and as well as novel peptides). Thus, in my view, many of the previous claims of identification of rare events in published proteogenomics studies, including recent high-profile *Nature* studies describing the first draft of the human proteome, need to be critically evaluated.

Thomas Kislinger: In theory any peptide sequence that is not present in a reference protein sequence database predicted from genomic/transcriptomic data and expressed abundantly enough to be detected by a modern shotgun proteomics strategy could be detected in a proteogenomics approach. This would include peptides with single-nucleotide variants (SNVs) and single-nucleotide polymorphisms (SNPs), insertions and deletions both in and out of frame, peptides that arise from aberrant splicing, and peptides that result from novel gene fusions. In addition, peptides that arise from translation of lncRNAs (long noncoding RNAs) or from intra- and intergenic regions of the genome could be identified by a proteogenomics approach. An additional caveat, aside from peptide concentration, might also be that some peptides are simply not amendable to mass spectrometric detection (i.e., biophysical properties).

Jacob Jaffe: In the future it should be the norm that every sample analyzed by proteomics is in reference to the genome (or genomes) of the biological sample being interrogated. It's not a question of what is missed by "classical" proteomics; it's just that we're doing a bad job in proteomics by performing our analyses in reference to an average predicted proteome that is not really suitable for most samples.

Andrei Drabovich: Such variant peptides include SNVs and missense mutations, fusion genes, truncated proteins, splicing isoforms, and peptides produced through translation of pseudogenes, UTRs, intergenic regions, or noncoding RNAs.

Which biological or clinical unmet needs can be addressed by proteogenomic technologies?

Alexey Nesvizhskii: Proteogenomics analysis can be useful as part of any study where more complete characterization of the genomic and proteomic diversity is desired. It is well established that that joint analysis of protein and mRNA data can provide biological insights not apparent from the analysis of each data type alone. Proteogenomics adds more depth to such integrative analyses by allowing detection and quantification of novel peptides that are expressed in a particular sample but would be missed when using a standard reference protein sequence database. From a proteomics perspective, for any biological or clinical question that can be addressed using proteomics and where relevant genomic data can be obtained (e.g., generated as part of that study or obtained from public sources), proteogenomics can provide an additional dimension. From the genomics/transcriptomics perspective, proteomics data can be extremely useful as a "proteomic filter," suggesting which of the genomics-based findings are more likely to be functionally significant because they propagate to the protein level.

Thomas Kislinger: Proteogenomics will continue to have an impact on genome annotation by providing direct evidence of what genes are ultimately translated to a detectable protein product. From a systems biology point of view, the integration of genomic, transcriptomic, and proteomic data will assist in our understanding of how information flows from gene to protein. Of course, whether such data will ultimately lead to a better mechanistic biological understanding or the identification of better clinical biomarkers is currently still highly speculative, since the field is still in its infancy. It will ultimately depend on what types of additional orthogonal information are gathered by using a proteogenomics pipeline. For example, it stands to reason that the relative abundance of a mutant peptide (and thereby its parent protein), compared to its native counterpart, could impact prognosis or response to treatment. So an unmet need is the characterization of peptides for important cancer genes that are suitable for targeted quantification, for example by multiple reaction monitoring–mass spectrometry (MRM-MS), and determining whether the ratio of mutated to endogenous (native) peptides in cancer can improve biomarker performance.

Jacob Jaffe: Proteogenomics really sets the stage for integrative biological analyses. For a long time "integration" has really just been a buzzword, but by aligning the genomics and the proteomics paradigms we can begin integrative analyses in earnest.

Andrei Drabovich: Regarding biological unmet needs, proteogenomic technologies may discover rare translational events in the cell, such as expression products of pseudogenes and noncoding RNA, classify protein isoforms and reveal functional impact of missense mutations, thus providing a rationale for drug discovery.

Unmet clinical needs to be addressed by proteogenomics would include development of personalized medicine approaches using patient-specific genomic and proteomic databases. This will facilitate more accurate stratification of cancer subtypes and may lead to more effective "personalized" therapy.

What is the potential impact of onco-proteogenomics in cancer research?

Alexey Nesvizhskii: Proteogenomics is very relevant to cancer research. Numerous published and ongoing efforts, e.g., the work done by The Cancer Genome Atlas (TCGA) consortium, employ genomics and transcriptomics technologies to generate in-depth profiles using cancer patient tissues as well as using cell line models. Data generated as part of these studies contain many sequence variants and novel transcripts that are potentially important for understanding the biological mechanisms of cancer progression or can be used as biomarkers for clinical diagnostics. An increasing number of cancer studies include proteomic analysis, exemplified by the efforts of the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) that performs proteomic profiling of TCGA samples. Also, it should be noted that the largest publicly available proteomics and genomics data sets that are suitable for proteogenomics analysis are coming from cancer-focused studies.

Thomas Kislinger: Again the impact of onco-proteogenomics on cancer research is currently highly speculative. The success and impact of onco-proteogenomics in cancer research will depend on several things. First, how many additional peptide sequences (and what types) can be confidently identified by generating custom proteogenomics databases. Secondly, what additional biological/clinical information (or utility) can be obtained through the identification of such peptide sequences? The most pressing issue is to determine if the identification (and quantification) of an onco-proteogenomic peptide can serve as a more sensitive biomarker, can provide additional information for the objective selection of better treatment modalities, or can be used to support novel biological insights. For example, while most genomic aberrations are likely to be neutral passenger mutations, onco-proteogenomics could assist in reducing the numbers by focusing on aberrations that are detectable and differentially regulated at the protein level.

Jacob Jaffe: I think onco-proteogenomics will allow for connecting-of-dots between cellular communication (phosphosignaling and other posttranslational modifications) and underlying genetic processes.

Andrei Drabovich: Onco-proteogenomics may complement genomics for stratification of cancer subtypes, discover molecular events upstream and downstream of known cancer biomarkers, and resolve the functional role of gene mutations, thus providing a rationale for drug discovery.

I would also speculate that some clinical tests based on the proteomic analyses of mutated peptides might find their unique niche in diagnostics of rare cancers and complement the next generation DNA sequencing. For instance, approximately 50% of glioma patients have a single missense mutation in the *IDH1* [isocitrate dehydrogenase 1 (NADP+)] gene (R132X, where X = G, H, or S). These 3 mutations can be measured by the targeted proteomic analysis of 3 corresponding tryptic peptides LVSGWVKPIIIG[X]^{132}HAYGDQYR. Isocitrate dehydrogenase 1 is a high-abundance intracellular protein, so proteomic assays could measure mutated IDH1 at the depth of 10^6 (an equivalent of one cancer cell detected in the presence of one million normal cells in the tissue biopsy). Since the deep next generation DNA sequencing currently provides the maximum depth of approximately 5×10^3, proteomic analysis of mutated peptides may facilitate the earlier diagnosis of some cancers. Diagnostic tests based on multiplex selected reaction monitoring (SRM) assays could target hundreds of mutated peptides and thus further increase diagnostic sensitivity, at 100% diagnostic specificity.

What are the major technical challenges in current proteogenomic pipelines?

Alexey Nesvizhskii: Proteogenomics requires high-quality proteomics data, ideally generated in parallel with transcriptomic data. This is not always feasible, and the number of high-quality, deep proteomic data sets suitable for proteogenomics studies is still limited. The biggest challenge, however, is bioinformatics. False peptide identification is a huge and underappreciated challenge of proteogenomics. Many published manuscripts, including high-profile publications in *Nature* and other journals, did not apply false discovery rate (FDR) assessment methods suitable for proteogenomics. In short, the stringency of filtering of novel peptides should be higher than that for known, commonly observed peptides. I have recently proposed a set of data analysis guidelines specifically for proteogenomics studies, and we are continuing to work in this area.

Thomas Kislinger: While it is certain that omics technologies have not yet reached their peak performance and can still be improved, I believe that the main challenges of proteogenomics are currently at the level of data analyses. This includes the development of appropriate statistical analysis frameworks to assign confidence values to the identification of proteogenomics peptides. In addition, one could even argue that we haven't even rigorously tested what is the most appropriate proteogenomics pipeline. For example, is an "individualized" approach combining NGS and shotgun proteomics on the same samples the best approach (and, as a subquestion: what type

of sequencing, WGS, exome-Seq, or RNA-Seq). Alternatively, a customized database using publically available variant (or mutant) peptide sequences [i.e., dbSNP (the Single Nucleotide Polymorphism Database), COSMIC (Catalogue Of Somatic Mutation In Cancer)] could be used. To the best of my knowledge this has not been rigorously tested to date and each strategy has its unique pros and cons. One final challenge to be addressed by proteogenomics is with regard to data deposition. Much like genomics, proteomics now has the capacity to detect germline SNPs which, in combination, have the potential to uniquely identify the patient. This raises potential ethical issues with regard to the deposition of raw data that will have to be resolved by the community.

Jacob Jaffe: We need to make it routine to assemble a customized proteomics database from any type of genomics/transcriptomics data. We further need to be able to take advantage of gene-level metadata in the context of these proteogenomic mappings. The lines between gene isoforms and proteoforms must disappear.

Andrei Drabovich: Sensitivity of mass spectrometers needs to be improved by an additional 4- to 5-fold to enable the quantitative analysis of proteins within the dynamic range of 10 orders of magnitude (the dynamic range of protein concentrations in mammalian cells and biological fluids). In addition, dedicated software packages for proteogenomic analysis and better statistical algorithms to control false-positive and false-negative rates of peptide matching will be required.

In what directions do you expect the field of proteogenomics to expand in the future?

Alexey Nesvizhskii: Proteogenomics has been around for several years, but its impact has been limited in part owing to low depth and low protein sequence coverage in data produced by the previous generations of proteomics technology. However, MS instrumentation has improved significantly and now allows deep proteome profiling, in some cases approaching the depth of RNA-Seq data. So one can start using proteogenomics, in a more meaningful way than before, to look for evidence of protein-level expression of transcripts currently annotated as noncoding RNAs or pseudogenes, to search for short ORFs and uORFs. The number of transcripts observed in any RNA-Seq data for any sample is astonishing. So the question on everybody's mind is which of these are actually translated? Which have any functional significance? There is a lot of interest in noncoding RNAs and uORFs, with questions about their protein coding potential; there is quite a bit of controversy as well. Proteogenomics can provide valuable information here. As ribosomal profiling technology improves, the combination of RNA-Seq, ribosomal profiling, and proteomics will provide a rich source of data for all sorts of proteogenomics studies.

Thomas Kislinger: Once the informatics challenges have been addressed and peptide false-discovery rates are accurately defined, including the selection of the most appropriate proteogenomics strategies, one possibility is that most cancer proteomics projects will apply some type of proteogenomic approach. The next logical steps would be to evaluate if detection of aberrant onco-proteogenomic peptides improves biomarker performance and to evaluate the functional relevance of such proteins in the context of cancer biology or therapeutic interventions. For example, does the expression of a mutated protein change its protein interaction partners, subcellular localization, or posttranslational modifications? In the context of biomarkers, one could develop targeted proteomics assays (i.e., MRM-MS) to specifically quantify onco-proteogenomic peptides. One additional expansion could be to include top-down proteomics, with its ability to accurately define specific proteoforms, in a proteogenomics pipeline. While this might not be technically feasible as of yet, this could ultimately provide the answer of how the cancer proteome is a reflection of the upstream cancer genome.

Jacob Jaffe: Pretty soon looking at a proteogenomics "track" in a genome browser (with quantitative information and posttranslational modifications mapped) will be routine. Next will be metaproteogenomics as more and more standardized data become available.

Andrei Drabovich: No doubt, there will be further integration of genomics, transcriptomics, and proteomics empowered by not only qualitative, but also accurate, quantitative information. Such integration will be later complemented by epigenetics, posttranslational protein modifications and metabolomics.

There is also a hope that proteogenomics will better stratify subtypes of cancers. Some very exciting discoveries related to the molecular mechanism of cancer could be made for some rare mutation-free cancers. For example, neither gene mutations nor epigenetic alterations were identified for posterior fossa type B ependymomas, for which cancer driver alterations may exist at the level of transcriptome, proteome or metabolome.

I also foresee that proteomic community will start actively using the ample genomic data generated by the large genomic consortiums such as the International Cancer Genome Consortium (ICGC), COSMIC, and TCGA. For example, the TCGA portal alone contains comprehensive genomic data, such as the whole exome sequencing, profiling of SNPs and DNA methylation and mRNA and miRNA analysis, for more than 11 000

individuals. Hopefully, these data will be translated into proteomic databases and drive further developments of proteogenomics.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: E.P. Diamandis, *Clinical Chemistry*, AACC.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: T. Kislinger, Prostate Cancer Canada; A.P. Drabovich, Movember Rising Star in Prostate Cancer Research grant.
Expert Testimony: None declared.
Patents: A.P. Drabovich, United States patent 9,040,464.

Previously published online at DOI: 10.1373/clinchem.2015.247858