Prioritizing Proteomics Assay Development for Clinical Translation (Additional Information)

SRM Assays

Selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM), is a mass spectrometry based assay commonly utilized by the proteomics community to detect and quantify proteins in a targeted and multiplexed manner. After an SRM assay is developed for a protein and validated in a sample, the assay can be easily disseminated digitally (e.g., as the set of instrument parameters) for broad adoption by other investigators and laboratories. Because SRM assays offer high specificity, accuracy and reproducibility, it is well suited for basic research objectives as well as biomarker discovery. SRM assays offer a number of important advantages over conventional Western blots, ELISA or other immunoassays. For example, (i) SRM assays can unequivocally differentiate and detect differences of as little as one amino acid residue between two protein isoforms or modified variants. Quantification assays can be designed to target a specific segment or modification site in a protein, which allow post-translational modifications sites and protein splice isoforms to be more easily utilized as potential biomarkers (e.g., cardiac troponins with site-specific phosphorylation). (ii) SRM assays are not restricted by the scarcity and high cost of high-quality antibodies. The development of an SRM assay requires significantly less time than the development of antibodies; hence SRM assays can be deployed quickly on multiple proteins to accelerate clinical translation. (iii) SRM assays have excellent multiplexing capabilities. Multiple proteins can be analyzed concurrently in a single assay, which effectively reduces batch-to-batch variations across replicates and samples. Finally, (iv): Isotope-tagged standards can be added to endogenous samples to allow for consistent and reproducible absolute quantification across multiple batches, sites and operators.

However, the adoption of SRM assays has occurred at a slow pace in both the clinical setting as well as in general basic research laboratories, and thus currently they remain the technical specialties of proteomics facilities and have limited opportunities to impact the broader research and biomedical community. Ideally one could envision that once an SRM method is developed by proteomics experts, it should be readily disseminated to benefit basic and clinical research in the broadest manner possible. As there are no distinctive logistical challenges in terms of information dissemination or material reagents, the current slow rate of adoption is thought to be due in part to insufficient efforts in developing SRM assays based on broad research and clinical needs, which occurs as a result of a disconnect between method developers and assay users, i.e., assays are either not developed for proteins of interest or the assays developed are not in high demand. Our data science approach has provided a solution for developers to determine which proteins are of the most interest to the clinical and basic research communities, thus allowing the prioritization of SRM assay development for highly demanded cardiac protein assays in both settings.

Availability of Targeted MS Assays

To identify existing targeted MS quantification resources for the list of popular proteins, we queried the SRMAtlas component of PeptideAtlas (1) that documents SRM transitions derived from shotgun proteomics experiments. As shown in the **Table** in this manuscript, we consider a targeted MS detection method to be available if any SRM transitions of a protein are documented on SRMAtlas (queried using its UniProt Accession). The majority of these SRM transitions were however inferred from shotgun proteomics experiments that have not been tested for protein detection by an actual SRM assay. We consider an experimental assay for a protein to be available if the protein is catalogued on the PeptideAtlas SRM Experiment Library (PASSEL) (2), which is a database cataloguing experiments that employed SRM assays to successfully detect a protein in a biological sample. It should be however noted that these SRM assays might have been developed for the biological samples that are not cardiac relevant (e.g., for an alternative protein isoform in the liver), and thus will require further development and optimizations of the assay in relevant samples and cardiovascular cohorts.

Functional Analysis

Gene Ontology analysis of the top mouse and human proteins was performed using WebGestalt and NCBI DAVID (3, 4). We further performed pathway analysis using Qiagen Ingenuity Pathway Analysis (IPA) on top human and mouse proteins in the heart (Figure S1). IPA protein network was overlaid on the popularity of a protein in mouse and human based on publication counts. The network diagram in Figure S1A shows that mouse proteins preferably included transcription factors and developmental proteins (e.g., NKX2-5, GATA4, TBX5), whereas human proteins are preferably involved in overt physiologic phenotypes (e.g., SCN5A). The functional differences between mouse and human proteins are corroborated in Figure S1B, which show the significance of enrichment of biological functional categories and pathways in the top mouse and human proteins. Mouse proteins were more highly enriched in cardiovascular system development categories.

Study Limitations

In this study, we proposed a data science method intended to rank the importance of a protein to cardiac research, as a surrogate of the importance of a protein to cardiac biology. We acknowledge that there are other methods and metrics to objectively evaluate the importance of a protein and thus the priority to which it should be assigned in designing protein quantification assays. However, when designing target assays to target the broadest audience, the popularity of a protein is an end unto itself: if a particular protein is intensively investigated by many, it would not be unreasonable to prioritize assay development for that protein such that the assays would be potentially adopted by broad interested parties.

Here we consider several other complementary metrics may also be employed to identify protein targets based on our evaluation of biological significance. In formulating our study design, we considered three additional possibilities. Firstly, gene networks may be constructed from co-expression or protein-protein interaction data to selectively target hub genes for protein quantification assays. The application of such assays to diverse biological samples may prove useful in unraveling regulatory networks. Secondly, we remark that the normalization of publication counts across multiple search terms may yield cardiac proteins that are more uniquely cardiac-relevant, and remove some protein (such as p53 and APOE) that are popular not only in cardiovascular sciences but also in other fields such as cancer and Alzheimer's research. Thirdly, it is possible non-objective metrics (e.g., based on aggregation of expert opinions or automated text-mining) may also prove fruitful and lead to additional insights in various decision-making processes.

We believe the above metrics are complementary in nature, and that the identification of high-frequency proteins are useful for certain applications, for instance in the initial development of protein assays that are meant to target the widest possible audience in multiple fields.

Pathway Analysis

Figure 1A

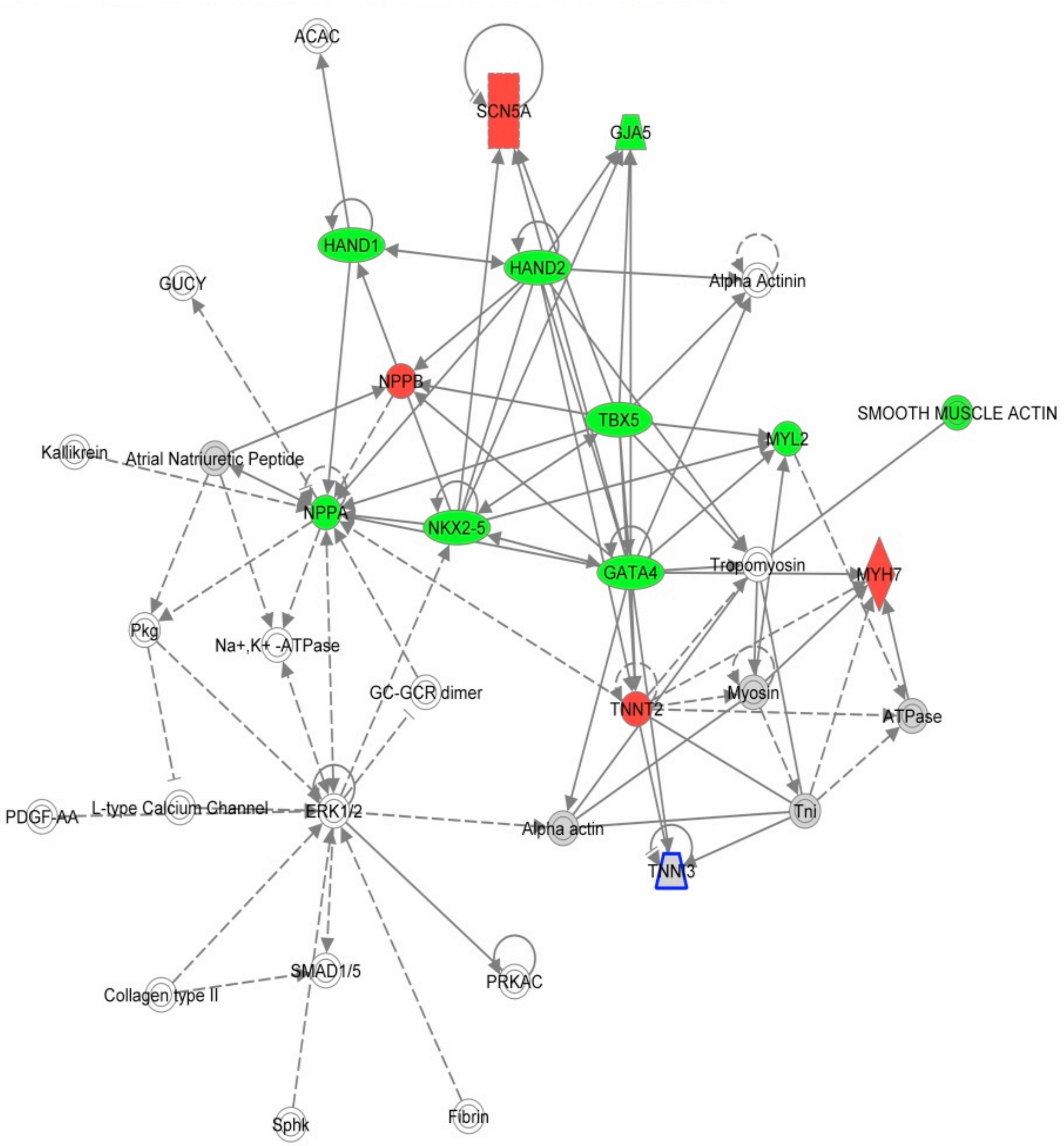

Figure 1B

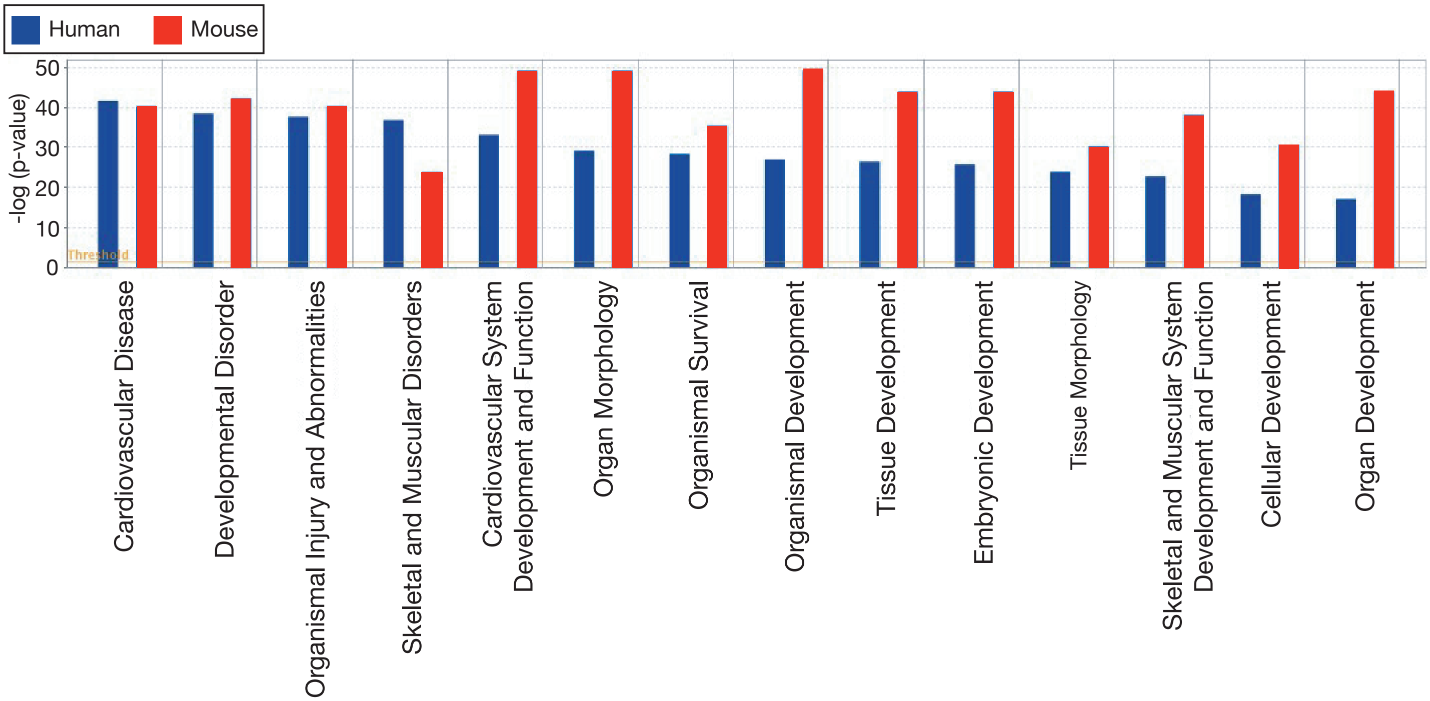

Figure 1. A. Protein networks of top mouse and human proteins. Green color of a protein node denotes a top-20 mouse protein; red color denotes a top-20 human protein; blue outline denotes a top-20 protein in both mouse and human (TNNI3). Protein networks of top mouse and human proteins illustrate the difference between most studied proteins in mouse and human. **B.** Ingenuity Statistical significance of the specific biological functions and pathways enriched in the human (blue) and mouse (red) datasets. Mouse proteins were more significantly enriched in developmental pathways.

References

1. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 2008;9:429–434.

2. Farrah T, Deutsch EW, Kreisberg R, et al. PASSEL: The PeptideAtlas SRMexperiment library. Proteomics 2012;12:1170–1175.

3. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 2013;41.

4. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.