

Title of proposed idea:

Significant U.S. Involvement in the Global Human Proteome Project—Sustainable Databases, Well-Annotated Parts Lists, Integrated Omics Studies, and Accelerated Clinical Applications of Proteomics

Nominator: Gil Omenn

University of Michigan

**What is the major obstacle/challenge in the field? What is needed to overcome this obstacle/challenge?**

Too many disparate, unconnected projects with limited sensitivity, transparency, and reproducibility. Need to identify, characterize, and quantitate post-translational modifications and alternative splice variants and their regulation. Need to address heterogeneity within tissue specimens and across patients with similar diagnoses.

A more coordinated framework, with stringent criteria for credible protein identification and confirmation, orthogonal validation, organized characterization of biological and disease-associated features, bold technological goals, and sustainable data resources linked to other omics resources.

**What emerging scientific opportunity is ripe for investment by the Common Fund?**

The European Union 7<sup>th</sup> Framework and multiple funders in Europe and Asia are investing substantially in critical components of the rapidly emerging global Human Proteome Project. The U.S., by contrast, has a broad array of mostly unconnected project studies with limited output of validated biomarker candidates and limited systems biology understanding.

The remarkable progress in mass spectrometry instrumentation and analytical strategies, in antibody and other protein capture for protein identification and tissue/organellar localization, and in well-curated, re-analyzed, inter-connected data resources makes a well-coordinated effort timely and likely to be very productive.

The U.S. could make a special contribution by, first, connecting mass spectrometry and protein capture approaches in biology and disease-driven major studies; second, connecting multiple omics levels to bridge the gap between genomes/gene regulation and phenomes; and, third, sustaining the critical U.S.-based data resources, namely PeptideAtlas and Tranche, both of which are critical to the global ProteomeXchange (with EU funding primarily for the European Bioinformatics Institute and NeXtProt components) and both of which are currently vulnerable due to ending of project-based support.

**What are the potential Common Fund investments that could accelerate scientific progress in this field?**

1. Multi-year support for critical data resources (about 3-4 FTEs each for PeptideAtlas and for Tranche) plus 2 FTEs for the HUPO Human Proteome Project web portal, with plans and requirements that NIH-supported studies be published with full transparency of the data and any novel algorithms to permit replication of the findings.
2. Coordinated multi-institute NIH information sharing and resource provision of SRM peptides and SRM Atlas, of current or expanded polyclonal and monoclonal antibody resources and

**related libraries of aptamers and other capture agents, with data submission to [www.antibodypedia.org](http://www.antibodypedia.org) of experience with available agents.**

3. New technologies with sufficient sensitivity and specificity to examine heterogeneous cellular populations, including single cells, and to distinguish and quantitate PTMs and splice variants.
4. Disease-oriented studies with multiple omics platforms and data resources for systems biology/networks investigations of disease processes and targets for prevention and therapy.
5. Out-of-the-box concepts and test development strategies for clinical biomarker candidates both for diagnosis and for guided therapy.

**If a Common Fund program on this topic achieved its objectives, what would be the impact?**

1. More rapid, more efficient progress in the field of proteomics and in linking genomes to phenomes, greatly enhancing the impact of genomics over the next 5-10 years.
2. New technologies—building on the recent experience of LTQ-Orbitrap and its several successors for more accurate shotgun analyses, the triple quad mass spectrometers that accelerate SRM targeted proteomics, the ETD for PTMs—with greater sensitivity and specificity, and with new kinds of capture agents using chimeric molecules, nano, or other cross-field strategies.
3. More biologically-driven and less statistically-driven development, confirmation, and validation of biomarker candidates with potential for more specific diagnoses, possibly earlier diagnoses, and surely better choices of therapy.
4. Much more useful data resources that empower individual investigators and research consortia to generate important hypotheses and accelerate their own focused research.

Gil Omenn

9/17/11 draft