

COPaKB: A proteomic knowledgebase for cardiac biology and medicine

Peipei Ping, Nobel Zong, Haomin Li, Mabel Chen, Rafael Jimenez, Tao Xu, Henning Hermjakob, Huilong Duan, John Yates III and Rolf Apweiler

UCLA David Geffen School of Medicine, USA

We engineered a Cardiac Organellar Protein Atlas Knowledgebase (COPaKB) to advance cardiovascular biology and medicine via proteomics. Orthogonal datasets on functional proteomics and cardiovascular biology are contextualized in a relational database, managed by MySQL. A modularly structured peptide spectral library serves as the scaffold of the COPaKB with a resolution of subcellular milieu. Four modules with 3,171 proteins are currently compiled; plateaus were reached for each module. Novel web services are implemented to disseminate proteomic proficiencies across the globe. As an example, large raw spectral files are dissected into small data packages and characterized with the COPaKB as references, in a reliability and reproducibility fashion. A benchmark test demonstrated the sensitivity of library mediated spectral analyses. The library search covered 83.9% of the proteins identified via a database search, as well as a bonus 51.6% at the same statistical cutoff. Furthermore, with raw spectral files as input, COPa knowledgebase automates targeted compilation of proteome properties of interests. This ideally bridges data-driven discoveries with hypothesis-driven investigations, and expedites in-depth analyses. To foster synergy among proteomic researchers, each user is invited to participate the restructuring of the core datasets with individually approved consent. Overall, the implementation of this specialized protein knowledgebase leverages state-of-the-art technology and curated datasets among the research community at large. COPaKB provides an avenue to unleash the full potentials of proteomics in the practices of cardiac biology and medicine.

peipeiping@earthlink.net