



Published in final edited form as:

Circ Genom Precis Med. 2019 April ; 12(4): e002426. doi:10.1161/CIRCGEN.118.002426.

HeartBioPortal: An Internet-of-Omics for Human Cardiovascular Disease Data

Bohdan B. Khomtchouk, PhD^{1,2,3}, Kasra A. Vand, BS^{4,*}, William C. Koehler, BS^{4,*}, Diem-Trang Tran, BS⁵, Kai Middlebrook, BS⁴, Shyam Sudhakaran, BS⁴, Christopher S. Nelson, PhD⁶, Or Gozani, MD, PhD¹, and Themistocles L. Assimes, MD, PhD^{2,3}

¹Department of Biology, Stanford University;

²Department of Medicine, Stanford University School of Medicine;

³VA Palo Alto Health Care System;

⁴Quiltomics, Palo Alto, CA;

⁵School of Computing, University of Utah, Salt Lake City, UT;

⁶Independent Researcher

Journal Subject Terms

Computational Biology; Genetics; Gene Expression and Regulation

Keywords

bioinformatics; cardiovascular disease; genetics; database

Cardiovascular disease (CVD) is the leading cause of death worldwide, responsible for over 17 million deaths annually, a rate which outpaces even that related to cancer. Despite these sobering statistics, the state-of-the-art in computational infrastructure for the study of contemporary datasets related to CVD lags substantially behind that widely available in oncology, where improved data science and visualization methods have delivered publicly available comprehensive cancer genomics resources like Memorial Sloan Kettering Cancer Center's cBioPortal^{1,2} and the National Cancer Institute's Genomic Data Commons (GDC) Portal^{3,4}. In our view, such portals do an outstanding job of transforming data from The Cancer Genome Atlas (TCGA) into logical data visualizations that provide additional biological insight. Developing a similar user-friendly computational platform for CVD could significantly lower the barriers of discovery by providing researchers with rapid, intuitive,

Correspondence: Bohdan Khomtchouk, PhD, Department of Medicine, Division of Cardiovascular Medicine, Stanford University, 1070 Arastradero Road, Palo Alto, CA 94304, Tel: 4155306713, bohdan@stanford.edu.

* contributed equally

Disclosures: The authors declare that all supporting data are available within the article and can be accessed at http://www.heartbiportal.com/user_api. Stanford University has filed OTL disclosure on the methods described, and BBK, OG and TLA are named as inventors. BBK is a co-founder of Quiltomics. OG is a co-founder of EpiCypher, Inc. and Athelas Therapeutics, Inc. The other authors report no conflicts.

and high-quality visual access to molecular profiles and clinical attributes from existing CVD projects.

To our knowledge, no open-access computational resource for interactive visual exploration of multidimensional -omics datasets focused on CVD exists that rivals the utility, simplicity, and power of cBioPortal or the GDC Portal. While a large amount of data from various CVD studies has been deposited in dbGaP, these data are only accessible to researchers after a time-consuming manual review of databases to identify relevant datasets and a somewhat onerous administrative data request process. Within the field of CVD research, the American Heart Association's Precision Medicine Platform (PMP) has significantly facilitated access to such data by streamlining the search, request, and transfer of controlled-access data and harmonizing datasets for cloud-based analyses⁵. Notwithstanding these important contributions of the PMP, we note that the Platform's output currently mimics the content of TCGA (i.e., raw and harmonized data with no preprocessing or visualization of its internals). While the strength of the PMP centers on providing curated CVD datasets, programming tools/tutorials, and forums for collaboration, a need that remains unmet is an informatics infrastructure that can quickly distill CVD-related datasets into easy-to-interpret, insightful figures and charts in a manner analogous to the cBioPortal and GDC Portal, but whose content is updated on a regular basis.

Here we present HeartBioPortal (<https://www.heartbiportal.com>), a publicly available web application that takes the first important steps towards fulfilling this need by integrating existing CVD-related omics datasets in real-time across the biological dataverse to provide intuitive visualization and analyses in addition to data downloads. The complementary focus of the HeartBioPortal platform is to provide community support for issuing gene/disease/variant-specific requests and visualizing the search results in a multi-omics context. By establishing an output like that of cBioPortal or the GDC Portal – containing preprocessed/analyzed data plus interactive visualizations accessible to the broader CVD and stroke research communities – we hope that HeartBioPortal significantly lowers the barriers between complex CVD datasets and researchers who require rapid, intuitive, and high-quality data visualization of molecular profiles and clinical attributes embedded within these datasets.

Currently, HeartBioPortal features gene expression, genetic association, and ancestry allele frequency information for 23606 human genes and 5792 variants across 12 broadly defined cardiovascular diseases spanning 199 research studies. A workflow architecture diagram is depicted (Figure). From a biological database perspective, HeartBioPortal currently syncs and harmonizes CVD-relevant data in real-time across at least the following publicly available resources: ClinVar, NHGRI-EBI GWAS catalog, OMIM, dbGaP, GTEX, CREEDS, HapMap, 1000Genomes, ExAC, TOPMed, gnomAD, Ensembl, and GEO. We coin the phrase “internet-of-omics” (an omics-centric hybrid of the internet-of-things (IoT) and internet-of-data fields) to describe the methods we've deployed to achieve such heterogeneous database interoperability. We believe the phrase appropriately emphasizes the challenges of extracting relevant biological information, in this case cardiovascular disease omics datasets, across various biomedical resource facilities (e.g., PubMed publications, metadata entries of data files stored in biological databases, etc.), where we have abstracted

the concept of “things” from hardware (devices) to software (databases). The overarching goal of our “internet-of-omics” algorithms is to establish massively integrated e-connectivity of omics datasets from multiple biological data silos linked to each other by communication networks based on metadata context. Harmonized tidy data downloads are provided directly through HeartBioPortal’s application programming interface (API) to support the data needs of academic clinical/laboratory CVD researchers while facilitating community transparency/cross-validation of the original data sources.

Future directions in the development of HeartBioPortal include substantially enriching the current offering of studies focused on gene expression by refining the phenotype definitions and integrating alternative splicing information (e.g., differential transcript usage and isoform-level expression) as well as adding CVD drug target identification and prioritization features to enable pre-competitive drug discovery by assisting early CVD candidate target decisions, thereby bridging cutting-edge computational biology with clinical research. We also plan to expand HeartBioPortal’s genetic association content (currently restricted to the NHGRI-EBI GWAS Catalog) to include diverse GWAS consortium efforts as incorporated in the Broad Institute’s Cardiovascular Disease and Cerebrovascular Disease Knowledge Portals. The end goal is to facilitate exploratory data analysis of integrative multi-omics CVD data spanning a diverse trove of expression, association, drug discovery, and population genetic information.

Acknowledgments:

BBK acknowledges and thanks the American Heart Association (AHA) for financial support through the AHA Postdoctoral Fellowship program. BBK also thanks Komal Vadnagara and Evan Kaeding for useful discussions.

Sources of Funding: Research reported in this publication was supported by the American Heart Association (AHA) Postdoctoral Fellowship grant #18POST34030375 (Khomtchouk).

References:

1. Cerami E, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. 2012; 2(5): 401–404. [PubMed: 22588877]
2. Gao J, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling*. 2013; 6(269): p11. [PubMed: 23550210]
3. Grossman RL, et al. Towards a Shared Vision for Cancer Genomic Data. *The New England Journal of Medicine*. 2016; 375: 1109–1112. [PubMed: 27653561]
4. Jensen MA, et al. The NCI Genomics Data Commons as an engine for precision medicine. *Blood*. 2017; 130: 453–459. [PubMed: 28600341]
5. Kass-Hout TA, et al. American Heart Association Precision Medicine Platform. *Circulation*. 2018; 137: 647–649. [PubMed: 29440193]

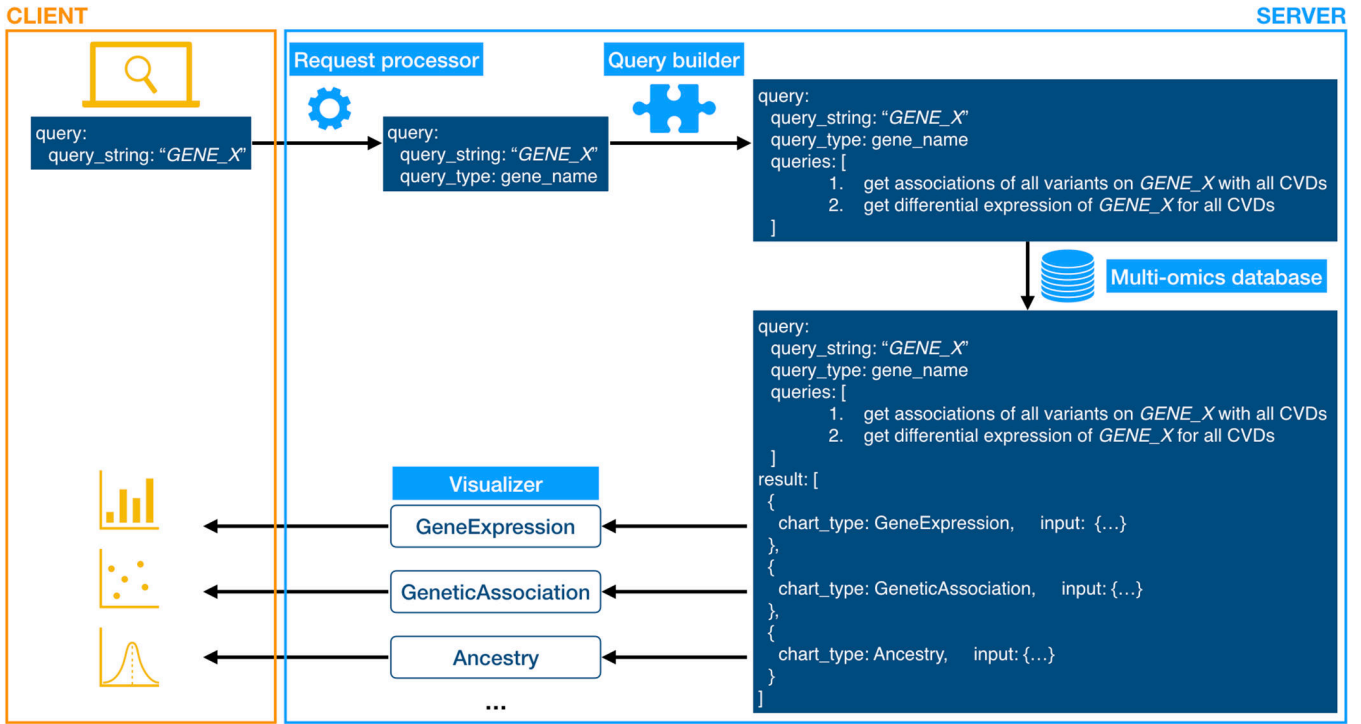


Figure. HeartBioPortal workflow architecture diagram. HeartBioPortal features a simple user interface that accepts queries for gene (current), disease/variant (future). User’s input is processed through multiple stages on the server into queries to be run on an integrated database, and eventually into preprocessed data that can be visualized in a meaningful way. Visualization components are highly extensible as the multi-omics database grows to include larger and more diverse sources of CVD data.