The Cardiovascular Disease Knowledge Portal: A community resource for cardiovascular disease research

Maria C. Costanzo, PhD[1]; Carolina Roselli, MS[2,3]; MacKenzie Brandes, MBA[1]; Marc Duby, MS[1]; Quy Hoang, BS[1]; Dongkeun Jang, MFA[1] Ryan Koesterer, BS[1]; Parul Kudtarkar, MS[4]; Annie Moriondo, MS[1]; Trang Nguyen, BS[1]; Oliver Ruebenacker, PhD[1]; Patrick Smadbeck, PhD[1]; Ying Sun, MS[4]; Adam S. Butterworth, PhD[5,6,7,8,9]; Krishna G. Aragam, MD, MS[10,11,12,13]; R. Thomas Lumbers, MD, PhD[14,15,16,17]; Amit V. Khera, MD, MSc[18,19]; Steven A. Lubitz, MD, MPH[12,20] Patrick T. Ellinor, MD, PhD[10,12,20]; Kyle J. Gaulton, PhD[4]; Jason Flannick, PhD[1,21,22]*; Noël P. Burtt, BS[1] **

Affiliations

1. Programs in Metabolism and Medical & Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2. Precision Cardiology Laboratory, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
3. Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands
4. Department of Pediatrics, University of California San Diego, La Jolla, CA, USA
5. BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

6. NIHR Blood and Transplant Research Unit in Donor Health and Behaviour, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
7. British Heart Foundation Cambridge Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge, UK
8. Victor Phillip Dahdaleh Heart & Lung Research Institute, University of Cambridge, Cambridge, UK
9. Health Data Research UK Cambridge, University of Cambridge, Cambridge, UK
10. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA.
11. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.
12. Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, MA, USA
13. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
14. British Heart Foundation Research Accelerator, University College London, London, UK.
15. Institute of Health Informatics, University College London, London, UK.
16. Health Data Research UK London, University College London, London, UK.
17. Bart's Heart Centre, St. Bartholomew's Hospital, London, UK.
18. Verve Therapeutics, Boston, MA, USA
19. Division of Cardiology, Brigham and Women’s Hospital, Boston, MA, USA
20. Demoulas Center for Cardiac Arrhythmias, Massachusetts General Hospital, Boston, MA, USA
21. Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA
22. Department of Pediatrics, Harvard Medical School, Boston, MA, USA

* Corresponding author:
Jason Flannick
Department of Pediatrics, Boston Children's Hospital, 300 Longwood Ave., Boston MA 02115
(617) 919-2670
jason.flannick@childrens.harvard.edu

** Corresponding author:
Noël P. Burtt
Director, Operations & Development, Diabetes Research & Knowledge Portals, Broad Institute of MIT and Harvard, Cambridge MA 02132
(617) 714-7635
burtt@broadinstitute.org

Short title: CVD Knowledge Portal

Word count: 800

Corresponding authors: Jason Flannick **(**jason.flannick@childrens.harvard.edu), Noël Burtt (burtt@broadinstitute.org)

Non-standard Abbreviations and Acronyms
CVD: cardiovascular disease

CVDKP: Cardiovascular Disease Knowledge Portal

HuGeAMP: Human Genetics Amplifier

HF: heart failure

CAD: coronary artery disease

AF: atrial fibrillation

Better understanding the mechanisms underlying cardiovascular disease (CVD) is critical for improving prediction, diagnosis, and treatment of these conditions affecting millions of people worldwide[1]. Large-scale genetic association studies have elucidated the genetic architecture of CVD biomarkers and outcomes, yet translating these data into biological knowledge is challenging. To assist with this challenge, we developed the Cardiovascular Disease Knowledge Portal (CVDKP; broadcvdi.org) in collaboration with CVD researchers. The CVDKP is built on the Human Genetics Amplifier (HuGeAMP)[2] and is a component of the Common Metabolic Diseases Knowledge Portal (cmdkp.org) along with portals focused on cerebrovascular disease[3], diabetes[2], and sleep disorders.

The CVDKP contains three dataset types. First, it includes 238 genetic association datasets for 173 phenotypes related to CVD, including results from the HERMES (heart failure, HF), CARDIoGRAMplusC4D (coronary artery disease, CAD), MiGen (myocardial infarction), AFGen (atrial fibrillation, AF), ICBP (blood pressure), GLGC (plasma lipids), and GIANT (anthropometric traits) consortia. In selecting genetic datasets, we prioritize: (a) CVD-relevant phenotypes not yet included in the CVDKP; (b) newer, larger datasets for major phenotypes; and (c) studies of non-European populations (to increase the diversity of ancestries represented in the

CVDKP). Second, the CVDKP contains >5,500 tissue-specific epigenomic annotations (*e.g.*, predicted cis-regulatory element locations), including nearly 350 for heart tissues and blood vessels. Finally, it includes lists of predicted "effector genes" for CAD, AF, HF, blood pressure, and plasma lipids, curated from publications that that prioritize causal genes at genetically associated loci. Although the CVDKP is funded by the Accelerating Medicines Partnership® in Common Metabolic Diseases (which includes pharmaceutical industry partners), decisions about datasets and content are made by academic CVD researchers and are not biased toward drug discovery. Datasets are biased, however, towards common rather than rare conditions due to the availability and focus of large-scale genetic association studies.

We apply a series of five bioinformatic methods (described fully in Costanzo *et al.*[2]) to these datasets to support high-level queries. We use a "bottom-line" meta-analysis method to compute a consensus p-value for a variant across all datasets, accounting for sample overlap; dataset-level associations are also viewable for each variant. This analysis eliminates associations in one dataset that are unsupported by other datasets (for example, 202 out of the 597 "dataset-level" associations across CAD, AF, and HF). It also uncovers novel associations that only become significant after meta-analysis (30 such associations across CAD, AF, and HF), including rs12209223 for AF, which lies within an intron of Filamin-A-Interacting Protein 1 (*FILIP1*), a gene expressed in skeletal muscle, cardiac tissue, and arteries. To highlight epigenomic annotations, pathways, or other traits related to a disease of interest, we apply MAGMA pathway analysis across all gene sets in the Molecular Signatures Database, stratified LD-Score regression across all epigenomic datasets, and cross-trait LD-Score regression across all pairs of traits. As illustrated for AF (**Figure**, panel A), these analyses together capture its genetic relationship to

disease causes (high BMI) and consequences (heart failure), underlying disease mechanisms (*e.g.*, atrioventricular blockage, cardiac muscle cell membrane depolarization), and tissues through which genetic associations exert their effects (*e.g.*, heart, cardiovascular system, muscle structure).

To access these results, the CVDKP contains four main pages, centered on a variant, gene, genomic region, or phenotype. Each page contains visualizations of associations and bioinformatic method results for the selected object and can be filtered to show ancestry-specific results when available. The portal also contains a menu of tools, each of which implements a multi-step filter or analysis. Some visualizations are intended for biologists unfamiliar with human genetics[4]: for example, the gene page includes a high-level summaries of a gene's "genetic support" (**Figure**, panels B and C), that omits technical details of association analyses. Other visualizations are intended for researchers well-versed in human genetics: for example, the tools menu includes a workflow that mimics widely used "variant to function" analyses (**Figure**, panel D) and programmatic access to all results. The CVDKP also draws information about gene product function and pathways from external resources: UniProt (https://www.uniprot.org/), MyGene (https://mygene.info/), and MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/).

The goal of the CVDKP is to provide a disease-specific resource that includes the genomic data types, datasets, and methods considered authoritative by the CVD research community. Its focus on genomic data complements other CVD-focused resources, such as HeartBioPortal[5] and BioDataCatalyst. We make the CVDKP accessible to all researchers – regardless of geographical location or career stage – by offering it as an open-access site with educational resources for all

learning styles, including written and video documentation, participatory webinars and focus groups, and a responsive helpdesk. We welcome new collaborations with the research community as we continue to develop the CVDKP.

All data and materials have been made publicly available at the CVDKP and can be accessed at https://broadcvdi.org/. The meta-analysis results from published GWAS datasets are not considered to be human subjects research and therefore do not require IRB approval.

Acknowledgements

We thank the many members of the CVD research community who have collaborated with us on populating and developing the CVDKP. In particular, Sekar Kathiresan was instrumental in launching the portal.

Sources of Funding

This work was supported by 2UM1DK105554 from the NIH. C.R. is supported by a grant from Bayer AG to the Broad Institute focused on the development of therapeutics for cardiovascular disease. P.T.E. is supported by grants from the NIH (1RO1HL092577, 1R01HL157635, 5R01HL139731), from the American Heart Association Strategically Focused Research Networks (18SFRN34110082), and from the European Union (MAESTRIA 965286). A.S.B was supported by core funding from the: British Heart Foundation (RG/13/13/30194; RG/18/13/33946) and National Institute for Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC-1215-20014; NIHR203312) [*]. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Disclosures

S.A.L. is a full-time employee of Novartis Institutes of BioMedical Research as of July 18, 2022. S.A.L. was previously supported by NIH grants R01HL139731 and R01HL157635, and American Heart Association 18SFRN34250007. S.A.L. received sponsored research support from Bristol Myers Squibb, Pfizer, Boehringer Ingelheim, Fitbit, Medtronic, Premier, and IBM, and has consulted for Bristol Myers Squibb, Pfizer, Blackstone Life Sciences, and Invitae. P.T.E. receives sponsored research support from Bayer AG, IBM Research, Bristol Myers Squibb and Pfizer; he has also served on advisory boards or consulted for Bayer AG, MyoKardia and Novartis. R.T.L receives sponsored research support from Pfizer. A.V.K. is an employee of Verve Therapeutics; has served as a scientific advisor to Amgen, Color Health, Foresite Labs, Illumina, Maze Pharmaceuticals, MedGenome, Navitor Pharmaceuticals, Novartis, Sanofi, Sarepta Therapeutics, Third Rock Therapeutics, and Veritas International; holds equity in Verve Therapeutics, Color Health, and Foresite Labs; and has received research funding from IBM Research. A.S.B. reports institutional grants outside of this work from AstraZeneca, Bayer, Biogen, BioMarin, Bioverativ, Novartis, Regeneron and Sanofi.

References

1. McClellan M, Brown N, Califf RM, Warner JJ. Call to Action: Urgent Challenges in Cardiovascular Disease: A Presidential Advisory From the American Heart Association. *Circulation*. 2019;139:e44-e54.

2. Costanzo MC, von Grotthuss M, Massung J, Jang D, Caulkins L, Koesterer R, Gilbert C, Welch RP, Kudtarkar P, Hoang Q, *et al*. The Type 2 Diabetes Knowledge Portal: an open access genetic resource dedicated to type 2 diabetes and related traits. *Cell Metabolism*. 2023, 35:P695-710.E6.

3. Crawford KM, Gallego-Fabrega C, Kourkoulis C, Miyares L, Marini S, Flannick J, Burtt NP, von Grotthuss M, Alexander B, Costanzo MC, *et al*. Cerebrovascular Disease Knowledge Portal: An Open-Access Data Resource to Accelerate Genomic Discoveries in Stroke. *Stroke*. 2018; 49:470-475.

4. Dornbos P, Singh P, Jang D, Mahajan A, Biddinger SB, Rotter JI, McCarthy MI, Flannick J. Evaluating human genetic support for hypothesized metabolic disease genes. *Cell Metabolism*. 2022;34:661-666.

5. Khomtchouk BB, Nelson CS, Vand KA, Palmisano S, Grossman RL. HeartBioPortal2.0: new developments and updates for genetic ancestry and cardiometabolic quantitative traits in diverse human populations. *Database (Oxford).* 2020;baaa115.

Figure legend

Scientific questions that can be addressed using the CVDKP. **A:** The Phenotype page, illustrated for the "Atrial fibrillation" (AF) phenotype, illuminates the genetic architecture of a disease or trait by displaying (top to bottom) a Manhattan plot of bottom-line genetic associations accompanied by a table (not shown) of the top associations; a plot and table of common variant gene-level associations and a table of pathways enriched for the top genes associated with AF, both determined using the MAGMA algorithm (de Leeuw CA *et al*., PLoS Comput Biol. 2015; 11:e1004219); and a table of phenotypes whose genetic architecture is correlated with that of AF, as determined by cross-trait LD-Score regression (Finucane HK *et al*., Nat Genet. 2015; 47:1228). **B**: Top, the Region page near the *MYO9B* gene shows significant associations for coronary artery disease (CAD), total cholesterol, and diastolic blood pressure. Bottom, scores from the Human Genetic Evidence calculator on the Gene page, which takes into account variant impact, common variant associations, and rare variant associations to generate summaries of the evidence for involvement of a gene in a disease or trait, predict that *MYO9B* may have roles in

CAD and traits that impact CAD risk. This gene has been experimentally shown to have a role in regulating vascular cell motility, supporting a role in CAD risk (Aragam K *et al*., Nat Genet. 2022; 54: 1803). **C:** Lists of predicted effector genes, curated from the literature, are available for multiple diseases and traits; the Gene page links to any effector lists on which that gene is represented. The *MYO9B* gene (see panel A) is included in 8 effector lists, for CAD, blood pressure, and lipid traits. Shown, part of the interactive table representing CAD effector gene predictions from the CARDIoGRAMplusC4D consortium (Aragam K *et al*., Nat Genet. 2022; 54:1803), which considered 8 types of evidence to predict the causal gene at each CAD-associated genetic locus. The table includes summary rows that are expandable (bottom) to show detailed evidence. **D.** The Variant Sifter displays genetic associations, credible sets, tissue-specific epigenomic annotations, and variant-gene links determined from chromatin conformation assays, as well as an interactive table (not shown) listing coordinates and parameters and linking to the source datasets. In this example, the LDL-associated variant rs2618566 near the *RRBP1* gene is seen to have high posterior probability in LDL credible sets, to be located in a predicted enhancer region in HepG2 cells, and to be linked to the RRBP1 promoter in both adipose and liver tissue, recapitulating findings by Ramdas *et al*. (Am J Hum Genet. 2022;109:1366).