Cardiovascular Disease Knowledge Portal

Costanzo, Maria C.; Roselli, Carolina; Brandes, Mackenzie; Duby, Marc; Hoang, Quy; Jang, Dongkeun; Koesterer, Ryan; Kudtarkar, Parul; Moriondo, Annie; Nguyen, Trang

Published in:
Circulation: Genomic and Precision Medicine

DOI:
10.1161/CIRCGEN.123.004181

Circulation: Genomic and Precision Medicine

RESEARCH LETTER

Cardiovascular Disease Knowledge Portal: A Community Resource for Cardiovascular Disease Research

Maria C. Costanzo, PhD; Carolina Roselli, MS; MacKenzie Brandes, MBA; Marc Duby, MS; Quy Hoang, BS; Dongkeun Jang, MFA; Ryan Koesterer, BS; Parul Kudtarkar, MS; Annie Moriondo, MS; Trang Nguyen, BS; Oliver Ruebenacker, PhD; Patrick Smadbeck, PhD; Ying Sun, MS; Adam S. Butterworth, PhD; Krishna G. Aragam, MD, MS; R. Thomas Lumbers, MD, PhD; Amit V. Khera, MD, MSc; Steven A. Lubitz, MD, MPH; Patrick T. Ellinor, MD, PhD; Kyle J. Gaulton, PhD; Jason Flannick, PhD; Noël P. Burtt, BS

Better understanding of 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[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.

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.

The CVDKP contains 3 dataset types. First, it includes 238 genetic association datasets for 173 phenotypes related to CVD, including results from the HERMES (Heart Failure Molecular Epidemiology for Therapeutic Targets), CARDIoGRAMplusC4D (coronary artery disease), MiGen (myocardial infarction), AFGen (atrial fibrillation [AF]), ICBP (International Consortium for Blood Pressure), GLGC (Global Lipids Genetics Consortium), and GIANT (Genetic Investigation of ANthropometric Traits) consortia. In selecting genetic datasets, we prioritize: (1) CVD-relevant phenotypes not yet included in the CVDKP; (2) newer, larger datasets for major phenotypes; and (3) studies of non-European populations (to increase the diversity of ancestries represented in the CVDKP). Second, the CVDKP contains >5500 tissue-specific epigenomic annotations (eg, predicted cis-regulatory element locations), including nearly 350 for heart tissues and blood vessels. Finally, it includes lists of predicted effector genes for coronary artery disease, AF, heart failure, 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, toward common rather than rare conditions due to the availability and focus of large-scale genetic association studies.

We apply a series of 5 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

Key Words: biomarkers ■ cardiovascular diseases ■ database ■ epigenomics ■ myocardial infarction ■ phenotype

Correspondence to: Jason Flannick, PhD, Department of Pediatrics, Operations & Development Boston Children's Hospital, Diabetes Research and Knowledge Portals, 300 Longwood Ave, Boston, MA 02115, Email jason.flannick@childrens.harvard.edu; or Noël P. Burtt, BS, Broad Institute of MIT and Harvard, Cambridge, MA 02132, Email burtt@broadinstitute.org

For Sources of Funding and Disclosures, see page 585 and 586.

Circulation: Genomic and Precision Medicine is available at www.ahajournals.org/journal/circgen

Nonstandard Abbreviations and Acronyms

AF	atrial fibrillation
CVD	cardiovascular disease
CVDKP	Cardiovascular Disease Knowledge Portal

variant across all datasets, accounting for sample overlap; dataset-level associations are also viewable for each variant. This analysis eliminates associations in 1 dataset that are unsupported by other datasets (eg, 202 out of the 597 dataset-level associations across coronary artery disease, AF, and heart failure). It also uncovers novel associations that only become significant after meta-analysis (30 such associations across coronary artery disease, AF, and heart failure), 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 [A]), these analyses together capture its genetic relationship to disease causes (high body mass index) and consequences (heart failure), underlying disease mechanisms (eg, atrioventricular blockage, cardiac muscle cell membrane depolarization), and tissues through which genetic associations exert their effects (eg, heart, cardiovascular system, muscle structure).

To access these results, the CVDKP contains 4 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 multistep 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 [B] and [C]), which 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 [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 geographic 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.

ARTICLE INFORMATION

Affiliations

Programs in Metabolism and Medical and Population Genetics (M.C.C., M.B., M.D., Q.H., D.J., R.K., A.M., T.N., O.R., P.S., J.F., N.P.B.), Precision Cardiology Laboratory (C.R.), Cardiovascular Disease Initiative (K.G.A., P.T.E.), and Program in Medical & Population Genetics (K.G.A.), The Broad Institute of MIT and Harvard, Cambridge, MA. Department of Cardiology, University of Groningen, University Medical Center Groningen, the Netherlands (C.R.). Department of Pediatrics, University of California San Diego, La Jolla (P.K., Y.S., K.J.G.). BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care (A.S.B.), NIHR Blood & Transplant Research Unit in Donor Health & Behaviour, Department of Public Health & Primary Care (A.S.B.), Victor Phillip Dahdaleh Heart & Lung Research Institute (A.S.B.), Health Data Research UK Cambridge (A.S.B.), University of Cambridge (A.S.B.). British Heart Foundation Cambridge Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, United Kingdom (A.S.B.). Cardiovascular Research Center (K.G.A., P.T.E.), Center for Genomic Medicine (K.G.A.), and Demoulas Center for Cardiac Arrhythmias (S.A.L., P.T.E.), Massachusetts General Hospital, Boston. British Heart Foundation Research Accelerator (R.T.L.), Institute of Health Informatics (R.T.L.), Health Data Rsrch UK London (R.T.L.), University College London. Bart's Heart Center, St. Bartholomew's Hospital, London, United Kingdom (R.T.L.). Verve Therapeutics (A.V.K.). Division of Cardiology, Brigham and Women's Hospital (A.V.K.). Department of Pediatrics, Boston Children's Hospital (J.F.). Department of Pediatrics, Harvard Medical School, Boston, MA (J.F.).

Acknowledgments

The authors thank the many members of the cardiovascular disease research community who have collaborated with us on populating and developing the cardiovascular disease knowledge portal. In particular, Sekar Kathiresan was instrumental in launching the portal.

Sources of Funding

This work was supported by 2UM1DK105554 from the National Institutes of Health (NIH). C. Roselli is supported by a grant from Bayer AG to the Broad Institute focused on the development of therapeutics for cardiovascular disease. Dr Ellinor 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). Dr Butterworth 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 Center (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

Dr Lubitz is a full-time employee of Novartis Institutes of BioMedical Research as of July 18, 2022. Dr Lubitz was previously supported by National Institutes of Health (NIH) grants R01HL139731 and R01HL157635, and American Heart Association 18SFRN34250007. Dr Lubitz 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. Dr Ellinor 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. Dr Thomas Lumbers receives sponsored research support from Pfizer. Dr Khera 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. Dr Butterworth 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. doi: 10.1161/CIR.0000000000000652
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 Metab*. 2023;35:695–710.E6. doi: 10.1016/j.cmet.2023.03.001
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. doi: 10.1161/STROKEAHA.117.018922
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 Metab*. 2022;34:661–666. doi: 10.1016/j.cmet.2022.03.011
5. Khomtchouk BB, Nelson CS, Vand KA, Palmisano S, Grossman RL. HeartBioPortal20: new developments and updates for genetic ancestry and cardiometabolic quantitative traits in diverse human populations. *Database (Oxford)*. 2020;2020:baaa115. doi: 10.1093/database/baaa115
6. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11:e1004219. doi: 10.1371/journal.pcbi.1004219
7. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228–1235. doi: 10.1038/ng.3404
8. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang M, Hindy G, Zhou W, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54:1803–1815. doi: 10.1038/s41588-022-01233-6
9. Ramdas S, Judd J, Graham SE, Kanoni S, Wang Y, Surakka I, Wenz B, Clarke SL, Chesi A, Wells A, et al. A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. *Am J Hum Genet*. 2022;109:1366–1387. doi: 10.1016/j.ajhg.2022.06.012

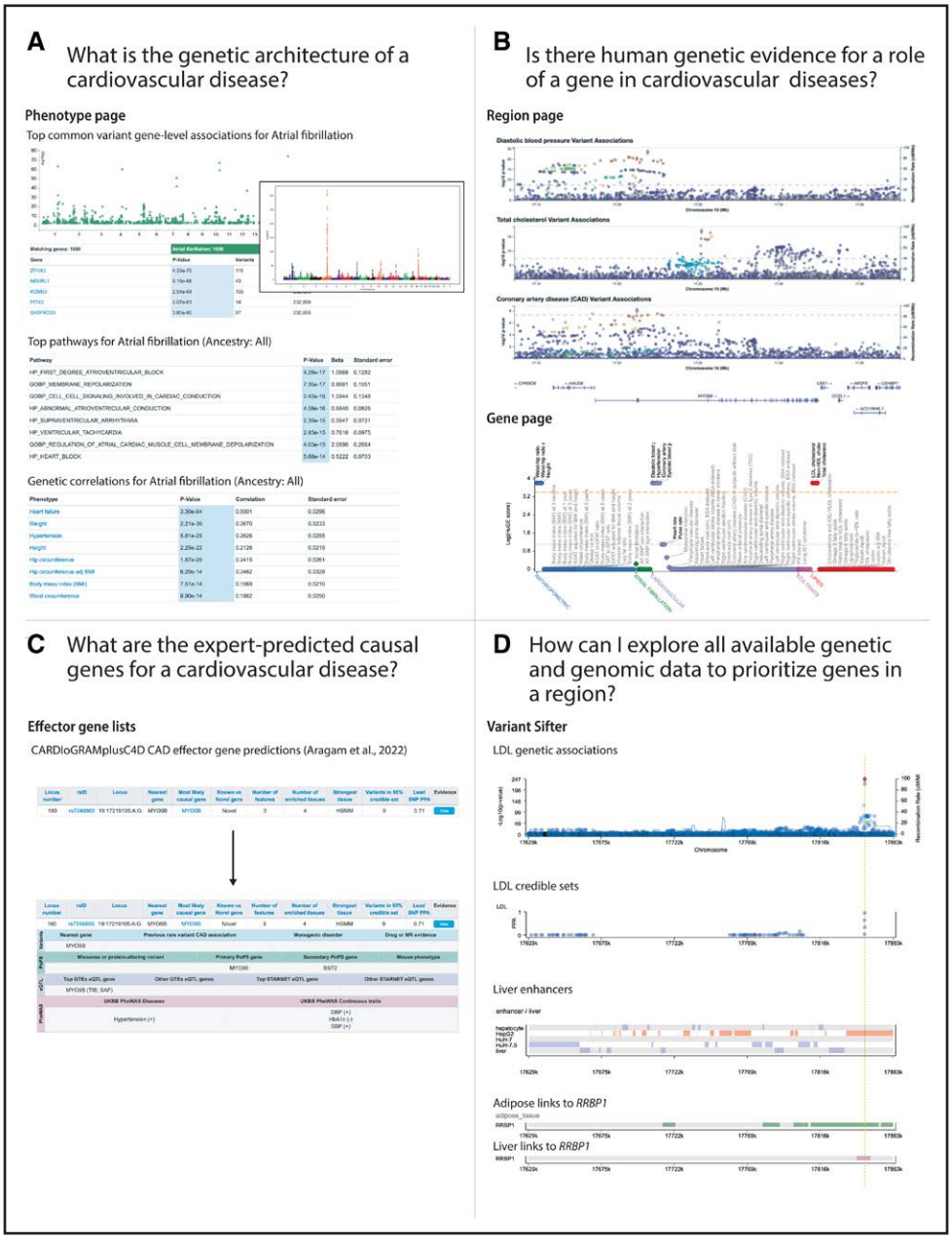

Figure. Scientific questions that can be addressed using the Cardiovascular Disease Knowledge Portal.
A, The phenotype page, illustrated for the atrial fibrillation (AF) phenotype, illuminates the genetic architecture of a disease or trait by displaying (top to bottom) a Manhattan plot of bottom-line 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[6]; and a table of phenotypes whose genetic architecture is correlated with that of AF, as determined by cross-trait LD-Score regression.[7] **B**, Top, the region page near the *MYO9B* (Myosin IXB) 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.[8] **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 (**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,[8] 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 (low-density lipoprotein)-associated variant rs2618566 near the *RRBP1* (Ribosome Binding Protein 1) 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.[9]