



Published in final edited form as:

Nat Genet. 2016 June 28; 48(7): 702–703. doi:10.1038/ng.3582.

Rapid evaluation of phenotype, SNP and summary results through the dbGaP Charge Summary site

Stephen S. Rich¹, Zeng Y. Wang², Anne Sturke², Lora Ziyabari², Mike Feolo², Christopher J. O'Donnell^{3,4}, Ken Rice⁴, Joshua C. Bis⁵, and Bruce M. Psaty⁶

¹Center for Public Health Genomics, University of Virginia, Charlottesville, VA

²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

³The Framingham Heart Study, Division of Intramural Research, National Heart, Lung and Blood Institute, Framingham, MA

⁴Division of Cardiology, Massachusetts General Hospital, Boston, MA

⁵Department of Biostatistics, University of Washington, Seattle, WA

⁶Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA

⁷Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, Health Services, University of Washington & Group Health Research Institute, Group Health Cooperative, Seattle, WA

The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium was established to conduct prospective meta-analyses for discovery of single nucleotide polymorphisms (SNPs) associated with phenotypes critical to heart, lung and blood disorders through genome-wide association studies (GWAS)^{1, 2}. The CHARGE Principles of Collaboration committed to broad sharing of GWAS results. It was reported, however, that public posting of summary results could, under certain conditions, be used to identify individuals in contributing studies^{3–5}. In response, the NIH moved large-scale GWAS summary results behind the IRB protection of the NCBI database of Genotype and Phenotype (dbGaP)⁶. Although p-values could be safely released, public posting of detailed GWAS results decreased dramatically. This decreased posting of GWAS results led to

Corresponding author: Stephen S. Rich, PhD, ssr4n@virginia.edu.

URLs. CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium Summary Results from Genomic Studies (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000930.v1.p1)

Accession codes. CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium Summary Results from Genomic Studies: phs000930

AUTHOR CONTRIBUTIONS

Conceived and designed the paper: SSR BMP. Developed and implemented the dbGaP site: ZYW, AS, LZ, MF. Coordinated analysis, transfer of GWAS data and deposition into dbGaP: CJO, KR, JCB, SSR. Wrote and edited the paper: SSR, ZYW, AS, LZ, MF, CJO, KR, JCB, BMP.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

investigators requesting “look-ups” of analytic results, a tedious process for both authors and investigators.

Here, we describe the CHARGE Summary Results page, the result of a joint project between the CHARGE consortium and dbGaP to allow public dissemination of GWAS summary statistics without the requirement for approval from a data access committee (DAC). Journals are now mandating deposition of results and data as part of the publication process. With an interest in creating a stable repository, CHARGE and dbGaP have developed a mechanism that protects participant confidentiality yet provides widespread data sharing through public posting of limited association information for each variant (See **Accession codes** for the dbGaP accession number for the CHARGE Summary Results page). The first public release occurred in winter 2015, with updates to be made quarterly.

The CHARGE consortium has published over 250 meta-analyses on a variety of phenotypes. These publications typically highlight only the findings that attain statistical significance. The vast majority of meta-analytic results, those failing to meet stringent statistical thresholds, do not appear in publications; however, many investigators desire to know whether “their” SNP (or set of SNPs), is associated with a phenotype at a reduced level of significance. The CHARGE Summary site provides information that can be viewed without Data Access Request (DAR) approval by a Data Access Committee DAC, providing for each phenotype and SNP rsID: (1) the “coding allele”, (2) the effect size (no direction) and (3) the p-value from the meta-analysis. In this way, based upon current dbGaP guidelines, genetic association data of interest to the scientific community is widely available while protecting the privacy of study participants. For those wanting more information, access to detailed summary statistics is possible; this requires DAR approval by the NHLBI DAC.

In order to facilitate the process of collection of the statistics by CHARGE investigators and protocol for transfer to dbGaP, the CHARGE wiki has been used to provide organizational and source information as well as example README files (Supplementary Note). As in standard prospective meta-analyses, an approved analytic plan was conducted separately within each study cohort. For each SNP, GWAS-specific results were combined using inverse-variance weighted meta-analysis. All studies conducted linear (or logistic or Cox) regression analysis using an additive genetic model. Background material on each study and contributing cohort information is provided on the open site (Supplementary Fig. 1).

With controlled access (DAR approval), the investigator obtains full results from association meta-analysis between genotype and phenotype across all analyzed SNPs with description of the study, analytic plan, and link to detailed results (Supplementary Fig. 2). The downloadable results also include the effect size (with direction of association), its 95% confidence interval, and the coded allele frequency for each SNP. The investigator can delve more deeply into the results and access additional genomic data resources by linking the analysis to association results in the Genome Browser (Supplementary Fig. 3). The initial view of the results⁷ in the Genome Browser (Supplementary Fig. 3) provides association results, color-coded in a heat map (red is $-\log(p) > 10$, or $p < 1 \times 10^{-10}$). A threshold determining which SNP results to visualize can be established with a pull-down tab, restricting the results to those with a specific value (e.g., $p < 1 \times 10^{-8}$). Interactive features of

the Genome Browser include the ability to explore results by movement of the mouse over the region (Supplementary Fig. 3, on the red area) to link to other components that provide views including other SNPs in that region as well as details of the individual SNP association results, gene structures and additional customizable tracks (Supplementary Fig. 4). Further exploration of annotations (*e.g.*, splicing) and links to other portals are available. These available results are summaries from meta-analysis; thus, no individual-level data from the studies are directly accessible, either from public or controlled access. In contrast, the genomic data-sharing beacons using multiple SNP queries of individual-level data have been implicated in privacy risks⁸. Indiscriminate use of this dbGaP site may result in publication bias, as investigators may fail to pursue publication if their specific SNPs do not attain support. However, these GWAS results are made available in order to better focus research without inhibiting publication of results. The use of the site can be reported as part of study methods, to provide readers with the source of the level of support from these large meta-analyses.

Providing access to study results should be encouraged, yet contributors of these summary results, and the supporting portal, cannot be responsible for potential abuse of the system. In summary, the CHARGE consortium has utilized meta-analysis to permit analysis of data across studies to adhere to limits of individual-level data sharing, an approach that improves statistical power to detect genomic regions associated with disease, biomarkers and risk factors. The development of the dbGaP CHARGE Summary Results site provides open access to the fundamental association results for a genotyped SNP and phenotype. With controlled access, an approved investigator has a full range of results and links to useful tools to enhance research. Open access to summary statistics in dbGaP provides detailed et non-identifiable research results. We recommend that other groups pursue similar collaboration with dbGaP to make summary data available from their GWAS data as well as other data types.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank all CHARGE investigators who have generated the analyses and manuscripts that constitute the summary statistics provided in the dbGaP portal. This work was generated without funding for the project. The CHARGE Consortium infrastructure grant to support meetings, working groups, and young investigators was provided to BMP (5 R01 HL105756) from the NIH. This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine.

References

1. Psaty BM, et al. *Circ Cardiovasc Genet.* 2009; 2:73–80. [PubMed: 20031568]
2. Bis JC, Glazer NL, Psaty BM. *J Thromb Haemost.* 2009; 7(Suppl 1):308–311. [PubMed: 19630823]
3. Homer N, et al. *PLoS Genet.* 2008; 4:e1000167. [PubMed: 18769715]
4. Visscher PM, Hill WG. *PLoS Genet.* 2009; 5:e1000628. [PubMed: 19798439]
5. Lumley T, Rice K. J. *Am Med Assoc.* 2010; 303:659–660.
6. Mailman MD, et al. *Nat Genet.* 2007; 39:1181–1186. [PubMed: 17898773]
7. Lemaitre RN, et al. *PLoS Genet.* 2011; 7:e1002193. [PubMed: 21829377]

8. Shringarpure SS, Bustamonte CD. *Am J Hum Genet.* 2015; 97:631–666. [PubMed: 26522470]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript