



Benefits of Cohort Studies in a Consortia-Dominated Landscape

Selam Zenebe-Gete, Rebecca Salowe and Joan M. O'Brien*

Department of Ophthalmology, Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA, United States

Keywords: GWAS - genome-wide association study, consortia, cohort, study design, genetics, big data

INTRODUCTION

Since their inception in 2005, Genome-Wide Association Studies (GWAS) have risen in popularity as a key method for gene discovery. Efforts to make use of existing genetic data, identify the causes of complex disease, and improve study power by increasing sample sizes have led to the formation of many GWAS consortia (Lutz et al., 2013). Consortia can bypass major logistical and financial challenges that accompany the recruitment of large study populations (Benjamin et al., 2018). Additionally, data sharing via consortia can help researchers diversify their study populations and can encourage research collaborations. While consortia do circumvent several financial, logistical, statistical, and demographic challenges to conducting GWAS, the collective swing towards their formation leaves several gaps in the field of genetics.

The majority of GWAS, including both small cohort studies and large consortia, often fail to identify actionable genetic determinants of complex disease. Only 2.2% of GWAS conducted between 2005 and 2016 had follow-up functional studies (Gallagher and Chen-Plotkin, 2018). Additionally, over 90% of phenotype-associated single nucleotide polymorphisms (SNPs) discovered in GWAS are in non-coding regions of the genome (Qu and Fang, 2013), with their effects on clinical outcomes either unknown or under investigation. Additionally, of the variants identified as statistically significant, a subset may not be clinically actionable because of their essential roles in cellular function. For example, p53 plays a significant role in many disease pathways (i.e. cancer), but its regulatory functions in all cells makes it a poor target for drug therapy.

Though these limitations apply to most GWAS, there are several unique advantages to cohort studies, including: 1) discovery of clinically actionable targets through subtyping; 2) the necessity of post-GWAS follow-up on variants; and 3) identification of population-specific findings.

SUBSECTIONS

Clinically Actionable Variants

Stratifying study populations by disease subtype may facilitate identification of clinically actionable variants. The discovery of variants through subtype-stratified GWAS has been well documented in the literature. In breast cancer research, stratified GWAS have identified subtype-defining SNPs, such as variants in the *HER2* gene (O'Brien et al., 2014). The development of therapies targeting *HER2* pathways improved outcomes for this subtype of breast cancer (Arteaga et al., 2012; Figueroa-Magalhães et al., 2014). Similarly, in an ischemic stroke study, all statistically significant variants were subtype-specific (Traylor et al., 2012). Subtype-stratified GWAS of bipolar disorder (BD) (Charney et al., 2017), autism spectrum disorder (ASD) (Hu et al., 2011), and Parkinson's disease (von Coelln and Shulman, 2016) patients also led to identification of novel disease subtype SNPs.

While subtyping can be implemented in both cohort and consortia studies, cohort studies may more easily accommodate subtyping methods, particularly when investigating complex diseases. Since some complex diseases do not have clearly defined subtypes, individual studies may define

OPEN ACCESS

Edited by:

Dov Greenbaum,
Yale University, United States

Reviewed by:

Chun Chieh Fan,
University of California, San Diego,
United States

*Correspondence:

Joan M. O'Brien
joan.o'brien@
pennmedicine.upenn.edu

Specialty section:

This article was submitted to
Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

Received: 25 October 2021

Accepted: 15 November 2021

Published: 07 December 2021

Citation:

Zenebe-Gete S, Salowe R and
O'Brien JM (2021) Benefits of Cohort
Studies in a Consortia-
Dominated Landscape.
Front. Genet. 12:801653.
doi: 10.3389/fgene.2021.801653

subtypes using different criteria. For example, a lack of methodological standardization in neuroimaging of Alzheimer's patients has hindered efforts towards consistent subtyping (Mohanty et al., 2020). Data harmonization methods are then needed to address differences in subtype classification between study cohorts within a consortium, increasing risk of introducing bias. A cohort study avoids this potential source of bias since it operates under a single set of subtyping criteria. Additionally, many cohort studies investigate a single disease type (i.e., open-angle glaucoma) rather than larger combinations (i.e., all forms of glaucoma), allowing for further subtyping of one form of disease. Since different forms of this disease may have different biological causes, putting them together to increase power may yield results that do not have biologic or physiologic relevance. Well-defined subtype-stratified GWAS in cohort studies may improve researchers' attempts to identify clinically actionable disease targets compared to consortia-based GWAS.

Post-Genome-Wide Association Study Follow-Up on Variants

Cohort studies tend to be more amenable to collecting longitudinal data and conducting follow-up on variants of interest (Wijmenga and Zhernakova, 2018). As more variants have been identified by GWAS, interest in investigating them further through functional studies, "multi-omic" research, and other "post-GWAS" methods has peaked. Cohort studies' focus on a single, typically local study population allows researchers to re-contact patients with interesting variants for follow-up. Cohort studies also allow for collection of longitudinal phenotypic data, which is key to the study of disease progression. Since consortia patients originally enrolled in studies at a variety of sites, patient re-contact and collection of longitudinal data are complicated and may not be possible in studies of such a wide scale. Furthermore, information collected at each site may lack consistency in data type and collection method, requiring additional harmonization efforts to make them comparable. For example, individual studies within the Psychiatric GWAS Consortium used different genotyping platforms, requiring imputation against existing gene expression data to standardize their genotypes (Sullivan, 2010). Even basic phenotypic parameters such as age and diet can affect an individual's gene expression (Wijmenga and Zhernakova, 2018), highlighting another difference between populations that can complicate combining multiple populations in consortia.

Population-Specific Variants

Cohort studies can also aid in identifying variants that are specific to minority populations, who remain dramatically under-represented in genetic studies. Population-specific GWAS are important to both understanding the genetics of complex traits and to elucidating the role of specific variants in minority populations (Sirugo et al., 2019). Studies have shown that differences in ancestry contribute to variations in disease prevalence, severity, and resistance across populations (Haga, 2010). However, while extensive GWAS testing has been done among individuals of European-descent, investigations of similar scale have not been conducted in African

ancestry populations (Campbell and Tishkoff, 2008), with only 3% of GWAS participants being of African descent as of 2016 (Popejoy and Fullerton, 2016).

Although some consortia can increase study population diversity through collaborative efforts to catalog human genetic variation, such as the 1000 Genomes Project (Jankovic et al., 2010), extensive data compilation may prevent identification of variants that are specific to underrepresented groups when they are un-stratified within the larger population. For example, an investigation of 3,899 SNPs in 313 genes in self-identified Caucasians, African Americans, Asians, and Hispanics found distinct and non-overlapping clustering of the Caucasian, African American and Asian samples (Stephens et al., 2001). This finding suggests differing genetic architecture between these groups, supporting the need for ancestry-specific genetic studies. Additionally, recruiting patients from a single city, as many cohorts do, minimizes the differences within a demographic group. Self-identified African Americans differ in their genetic admixture across different geographic locations (Bryc et al., 2015), suggesting that single city studies have less population heterogeneity than consortia.

DISCUSSION

While consortia do play an important role in genetic research, cohort-based studies may be better suited to identifying clinically actionable disease pathways and studying underrepresented minority populations. Although the recent publication climate emphasizes large consortia, GWAS of specified cohorts may produce more precise results that can be used in studies aiming to link genetics to endophenotypic data over time. Returning to the example of breast cancer, underlying pathology and genetics are now used to subtype patients in order to personalize treatment. One consortium including over 120,000 breast cancer patients identified 65 novel loci associated with overall breast cancer risk via GWAS (Michailidou et al., 2017). However, advancements in breast cancer technology have demonstrated the importance of molecular subtyping in patient prognosis and treatment (Yang and Polley, 2019). When multiple diseases or subtypes with distinct molecular pathways are categorized together in GWAS, as in the above study, the resulting genetic findings may not be clinically actionable. Putting this into context, future genetic studies may benefit from refocusing on the end goal of all GWAS—not only to find statistical significance, but to identify variants with the potential to improve health outcomes. Improving the accuracy of GWAS findings and translating these results to the clinic may be facilitated through a greater balance between both consortia- and cohort-based methods.

AUTHOR CONTRIBUTIONS

JO'B conceptualized the presented idea. SZ-G investigated and wrote the original draft. All authors reviewed and edited later drafts. All authors reviewed and approved the final version of the manuscript.

FUNDING

This work was supported by the National Eye Institute, Bethesda, Maryland (grant #1RO1EY023557-01) and Vision Research Core Grant (P30 EY001583). Funds also come from the F. M. Kirby

Foundation, Research to Prevent Blindness, The UPenn Hospital Board of Women Visitors, The Paul and Evanina Bell Mackall Foundation Trust, and Regeneron Pharmaceuticals. The Ophthalmology Department at the Perelman School of Medicine and the VA Hospital in Philadelphia, PA also provided support.

REFERENCES

- Artega, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F., and Gianni, L. (2012). Treatment of HER2-Positive Breast Cancer: Current Status and Future Perspectives. *Nat. Rev. Clin. Oncol.* 9 (1), 16–32. doi:10.1038/nrclinonc.2011.177
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2018). Redefine Statistical Significance. *Nat. Hum. Behav.* 2 (1), 6–10. doi:10.1038/s41562-017-0189-z
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., and Mountain, J. L. (2015). The Genetic Ancestry of African Americans, Latinos, and European Americans Across the United States. *Am. J. Hum. Genet.* 96 (1), 37–53. doi:10.1016/j.ajhg.2014.11.010
- Campbell, M. C., and Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genom. Hum. Genet.* 9, 403–433. doi:10.1146/annurev.genom.9.081307.164258
- Charney, A. W., Ruderfer, D. M., Stahl, E. A., Moran, J. L., Chambert, K., Belliveau, R. A., et al. (2017). Evidence for Genetic Heterogeneity Between Clinical Subtypes of Bipolar Disorder. *Transl. Psychiatry* 7 (1), e993. doi:10.1038/tp.2016.242
- Figueroa-Magalhães, M. C., Jelovac, D., Connolly, R. M., and Wolff, A. C. (2014). Treatment of HER2-Positive Breast Cancer. *Breast* 23 (2), 128–136. doi:10.1016/j.breast.2013.11.011
- Gallagher, M. D., and Chen-Plotkin, A. S. (2018). The post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* 102 (5), 717–730. doi:10.1016/j.ajhg.2018.04.002
- Haga, S. B. (2010). Impact of Limited Population Diversity of Genome-Wide Association Studies. *Genet. Med.* 12 (2), 81–84. doi:10.1097/gim.0b013e3181ca2bbf
- Hu, V. W., Addington, A., and Hyman, A. (2011). Novel Autism Subtype-Dependent Genetic Variants are Revealed by Quantitative Trait and Subphenotype Association Analyses of Published GWAS Data. *PLoS One* 6 (4), e19067. doi:10.1371/journal.pone.0019067
- Lutz, S. M., Fingerlin, T., and Fardo, D. W. (2013). Statistical Approaches to Combine Genetic Association Data. *J. Biomet. Biostat.* 04 (3), 1000166. doi:10.4172/2155-6180.1000166
- Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Jui, S., and Kar, S. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551 (7678), 92–94. doi:10.1038/nature24284
- Mohanty, R., Mårtensson, G., Poulakis, K., Muehlboeck, J.-S., Rodriguez-Vieitez, E., Chiotis, K., et al. (2020). Comparison of Subtyping Methods for Neuroimaging Studies in Alzheimer's Disease: A Call for Harmonization. *Brain Commun.* 2 (2), fcaa192. doi:10.1093/braincomms/fcaa192
- O'Brien, K. M., Cole, S. R., Engel, L. S., Bensen, J. T., Poole, C., Herring, A. H., et al. (2014). Breast Cancer Subtypes and Previously Established Genetic Risk Factors: A Bayesian Approach. *Cancer Epidemiol. Biomarkers Prev.* 23 (1), 84–97. doi:10.1158/1055-9965.EPI-13-0463
- Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is Failing on Diversity. *Nature* 538 (7624), 161–164. doi:10.1038/538161a
- Qu, H., and Fang, X. (2013). A Brief Review on the Human Encyclopedia of DNA Elements (ENCODE) Project. *Genomics Proteomics Bioinformatics* 11 (3), 135–141. doi:10.1016/j.gpb.2013.05.001
- Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I., and Boehnke, M. (2010). Genome-Wide Association Studies in Diverse Populations. *Nat. Rev. Genet.* 11 (5), 356–366. doi:10.1038/nrg2760
- Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177 (1), 26–31. doi:10.1016/j.cell.2019.02.048
- Stephens, J. C., Schneider, J. A., Tanguay, D. A., Choi, J., Acharya, T., Stanley, S. E., et al. (2001). Haplotype Variation and Linkage Disequilibrium in 313 Human Genes. *Science* 293, 489–493. doi:10.1126/science.1059431
- Sullivan, P. F. (2010). The Psychiatric GWAS Consortium: Big Science Comes to Psychiatry. *Neuron* 68 (2), 182–186. doi:10.1016/j.neuron.2010.10.003
- Traylor, M., Farrall, M., Holliday, E. G., Sudlow, C., Hopewell, J. C., Cheng, Y.-C., et al. (2012). Genetic Risk Factors for Ischaemic Stroke and its Subtypes (The METASTROKE Collaboration): A Meta-Analysis of Genome-Wide Association Studies. *Lancet Neurol.* 11 (11), 951–962. doi:10.1016/S1474-4422(12)70234-X
- von Coelln, R., and Shulman, L. M. (2016). Clinical Subtypes and Genetic Heterogeneity: Of Lumping and Splitting in Parkinson Disease. *Curr. Opin. Neurol.* 29 (6), 727–734. doi:10.1097/WCO.0000000000000384
- Wijmenga, C., and Zernakova, A. (2018). The Importance of Cohort Studies in the Post-GWAS Era. *Nat. Genet.* 50 (3), 322–328. doi:10.1038/s41588-018-0066-3
- Yang, S. X., and Polley, E. C. (2019). Systemic Treatment and Radiotherapy, Breast Cancer Subtypes, and Survival After Long-Term Clinical Follow-Up. *Breast Cancer Res. Treat.* 175 (2), 287–295. doi:10.1007/s10549-019-05142-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zenebe-Gete, Salowe and O'Brien. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.