Published in final edited form as:

Nat Genet. ; 44(2): 113. doi:10.1038/ng.1052.

THE AUTHORSHIP NETWORK OF GENOME-WIDE ASSOCIATION STUDIES

Brendan K. Bulik-Sullivan¹ and **Patrick F. Sullivan, MD FRANZCP²**

¹University of California, Berkeley, CA, USA

²Department of Genetics, University of North Carolina at Chapel Hill, NC, USA

Keywords

genome-wide association; genetics; human disease; network graph

Genome-wide association studies (GWAS) of complex human traits have become a major approach in human genetics. Taken together, GWAS are arguably the largest biological investigations of humans ever conducted. The number of people genotyped with a GWAS array is difficult to know, but probably exceeds 1,000,000. Major findings from these studies are that many common diseases have a polygenic architecture, the genetic effect sizes of common SNP variants are unexpectedly small, the identification of the involvement of genes and biological processes not previously suspected, and the association of some loci with different diseases.^{1–2} Critically, the sample sizes necessary to identify robust and replicable findings were beyond that achievable by single groups, and collaborations rapidly evolved to augment statistical power.

We sought to describe the collaborative networks that emerged as part of GWAS. Full methods are in the Supplemental Material. Briefly, we used the NHGRI GWAS catalog¹ and PubMed to identify the authors on 604 GWAS published from the first report in 2005 up to the last complete year (2010). These 604 GWAS papers had a total of 21,007 authorships (8,718 individuals).

We constructed network diagrams – graphs – where nodes are authors and edges connect co-authors on a GWAS paper. Overall, there was >10× increase in the number of co-authorships from 2005 to 2010. A network graph created using GEPHI is shown in **Figure 1**. The graph shows modularity at several levels: there are 14 empirical co-authorship modules (groups of nodes of the same color), and several modules show substructure (clusters within a module), and there are often abundant connections between modules. This graph is annotated more fully in **Figure S5** and **Table S6**, and is coherent in the identification of individuals, labs, and phenotypes studied.

Small-scale collaborations between laboratories are common in human genetics. There are multiple examples of large-scale, “big science” collaborations (e.g., the Human Genome

Correspond with Dr Sullivan: Department of Genetics, CB#7264, 5097 Genomic Medicine, University of North Carolina, Chapel Hill, NC, 27599-7264, USA. Voice: +919-966-3358, FAX: +919-966-3630, pfsulliv@med.unc.edu.

Project, HapMap, the 1000 Genomes Project, and ENCODE). Most large collaborations developed in a federated manner with a formal, top-down structure. GWAS collaborations are an unusual event in the history of biomedicine as large and extensive collaborations self-organized, and emerged rapidly from grassroots origins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367. [PubMed: 19474294]
2. Lander ES. Initial impact of the sequencing of the human genome. *Nature*. 470:187–197. [PubMed: 21307931]