The Rise of Consortia

Siiri Bennett and Neil Caporaso

Introduction

The collective accomplishments of genome wide association (GWA) studies have required the use of expanded data derived from consortia. GWAS have identified genetic loci reliably associated with diverse medical conditions and phenotypes, but less noted is the fact that these studies achieved something that two decades of candidate gene studies failed to do: assemble over 39 studies, each comprising over 50,000 participants (http://www.genome.gov/gwastudies/). A fundamental pillar of genomic era study design is that adequate control of type 2 error mandates lower bounds on sample size to reliably identify realistic risks (i.e. OR ~ 1.2-1.3) observed for common alleles[1-2]. Single institution's limited ability to accrue the necessary numbers of participants and the scarcity of genotyping resources have stimulated vigorous collaborative efforts. The increasing availability to investigators of in silico (dbGaP) data has played an important supportive role[3], but limits on phenotypic information, storage capacity, and rapidly changing technology constrain the benefits of this approach.

Sample size increases alone will contribute only marginally to revealing substantial additional components of moderate effect to as yet unidentified inherited contributions to common diseases, suggesting that enhanced technologies and refined understanding of genetic and phenotypic architecture will be needed[4-5]. However, investigators view identification of gene-environment (G*E) and gene-gene (G*G) effects and the exploration of key subgroups as active frontiers in genomics, with the attendant requirements for increased power/study size for these more demanding study goals[6]. These considerations all suggest that team science as facilitated by consortia will be crucial in the next wave of genomic discovery. In order to characterize this growing trend, we conducted two surveys and looked at the NHGRI catalog of published GWAS to explore the phenomenon of networking in genomics. We point to practices that appear to enhance resource utilization and scientific output and to facilitate communication and collaborations between GWAS investigators.

Surveys

In April 2011 we initiated two surveys (see **Supplemental Methods,** GWAS Investigator and Consortia Surveys). The first survey targeted all corresponding authors whose GWAS results were listed in the NHGRI compendium of published GWA studies (http://www.genome.gov/gwastudies/) through March 2011. A second survey was sent to representatives of consortia identified from public sources such as the National Cancer Institute (NCI) Epidemiology and Genetics Research Program (EGRP) list of consortia (http://epi.grants.cancer.gov/Consortia/), personal contacts, and searches based on Consortia web sites. The resulting data can be made available to interested investigators. A summary list of consortia from all sources is provided in **Supplemental Table 1**. Descriptive analyses were performed using Epi Info version 3.5.1 and Excel 2010. We also performed descriptive analyses of the data available in the NHGRI GWAS publications database, though the available information does not clearly distinguish

between those studies deriving from ad hoc collaborations formed around a specific analysis and those with a more formal organization.

Overall, in recent years there has been an increase in the median size of GWAS (**Figure 1**). Since individual contributing studies are relatively fixed by the size of participating study centers, increased total numbers of study participants reflect greater aggregations of studies, i.e. consortia (**Figure 2**).

Though information on all consortia was not available, the data still show some interesting trends. Consortia have investigated a diverse distribution of disease phenotypes, with cancer the most common (**Figure 3**). However, GWAS have a wealth of secondary phenotype data available (**Figure 4**), and in fact the largest GWAS studies listed in the NHGRI database of published GWAS are reports of analyses of secondary phenotypes such as smoking and alcohol consumption[7] (**Table 1**). While heritability for these secondary phenotypes is strong, challenges in phenotype definition (often primary studies were not designed to capture these detailed phenotypes) and differences in the genetic architecture of these traits likely contribute to lower magnitude of observed effect sizes. Funding of consortia is more likely for modest aggregations of participating centers (**Supplemental Figure 1**), and among those reporting funding, the NIH is the most common specific source (**Table 2**). Where racial group has been identified, more consortia have studied Caucasian subjects, though Asian and African participation is apparent across time and size indices (**Supplemental Figure 2**) and reflects the central importance of broad ethnic representation to genetic discovery[8] and global health[9-10]. Other trends in demographics (age group and gender), technology and analytic approach are described in **Supplemental Figures 3, 4, 5 and 6** respectively.

Future

A central question is what is the most effective way to promote the rapid conduct of high quality collaborative studies that require large numbers of subjects and therefore participation of a large number of groups? It is possible to establish very large new cohorts but the resources required to accumulate sufficient data will be quite substantial and many years required before sufficient numbers of endpoints have accumulated for meaningful analyses[11]. In contrast, minimal or modest additional support is needed to enable existing studies that have cataloged these outcomes to conduct collaborative population studies of size and scale.

Collaborations between GWAS investigators range from ad hoc agreements to conduct combined analyses of a specific phenotype (often supplemented by in silico warehouses like dbGaP), along a continuum to large, formalized and relatively permanent aggregations of studies supported by independent coordinating centers, steering committees, and explicit grant support[12]. An advantage of formal consortia is that they may serve as breeding grounds for collaborations with investigators outside the established framework who require even larger sample sizes to achieve the necessary statistical power or to identify subgroups. Funding consortia directly can enhance interdisciplinary participation, consistency in data quality, attention to issues of study design and harmonization[13] and rapid dissemination of 'best practices', enhance quality control and provide favorable economics[12] (**Table 3**). Such funding may provide a means to accelerate GWAS applications to important but less studied areas

such as pharmacogenomics[14], risk-models[15], or understudied populations. Although the size of a study alone does not necessarily correspond to the noteworthiness of the findings, even a null study of convincing size (i.e. power) can redirect scientific inquiry in important ways. But study size is only one reason to support more formal support of consortia. Formally supported consortia such as the Wellcome Trust[16], GAIN[17] and GENEVA[9] have helped develop and standardize practices that have contributed to successful GWAS methodologies and served as crucibles for secondary phenotype analyses that often include diverse collaborators beyond the confines of the original structure[18]. Initiatives to link and extend existing biorepository[19], genomic[20] and phenotype[21-22] resources are providing powerful new tools to extend GWAS discovery and application.

Moore's law[23] (the exponential improvement over time in processing speed and memory capacity) suggests that favorable trends in DNA analysis, information processing and storage will continue to accelerate. Also implied is the rapid obsolescence of today's technologies suggesting that support must be adaptive to new modes of collecting, distributing and analyzing genomic information. The expectations that successive waves of new technology will drive progress and that substantial sample sizes will be required to detect realistic population effects are safe ones but modes of data collection, assembly and analysis will rapidly evolve. Biosensors[24], personal devices, and social media will drive individual data collection in novel ways[25]. Nevertheless, consortia in some form will remain the major mechanism for obtaining sufficient numbers of most even moderately uncommon end points (e.g., non smokers with lung cancer). Given the importance of maximizing the impact of scarce resources, reducing the administrative burden of data-sharing while respecting ethical concerns is a critical priority[26]. While directly funding consortia is one straightforward approach, supporting critical infrastructures that enable networked initiatives involving repositories, databases, phenotype characterization resources (e.g., PhenX, https://www.phenxtoolkit.org) and tissue protocols are attractive strategies. Funding entities can enhance the public benefit by rewarding public posting of data and tools.

Team science is clearly responsible for an increasing proportion of breakthroughs in diverse fields, as demonstrated by the hundreds of GWA studies that have successfully identified genetic loci reliably associated with common complex diseases[27]. More complete data on studies made possible by consortia and greater attention to how their unique contributions can be facilitated are needed as the role of consortia is certain to be important going forward. Meanwhile, investigators can reap the immediate practical benefits of consortia and enhance their present collaborations by consulting and hopefully contributing to improvements of the posted catalog of consortia at http://www.wikigenes.org/GWAS/consortia.html . We suggest that investigators participating in consortia check their consortium listing, contribute any missing or erroneous data, and identify any new, emerging and evolving consortia (e.g., those undergoing name changes, splits and consolidations). A continuously updated listing of active consortia will be a useful resource to the genomics community for those seeking collaboration or information on publically available data involving a specific condition or phenotype. Review of these listings will be valuable to investigators at all stages of GWAS, but especially in the planning and replication phases of population studies.

Affiliations

Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, US Department of Health and Human Services, Rockville, Maryland, USA.
Neil E. Caporaso

Department of Biostatistics, University of Washington, Seattle, Washington, USA.
Siiri N. Bennett

Contributions

S.B. and N.C. developed the statistical approach and designed the analyses. S.B. implemented the methods. N.C. and S.B. drafted the manuscript, critically reviewing the paper, and approved the final version of the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Correspondence to:

Neil E. Caporaso

Acknowledgements

The authors wish to thank Teri Manolio of NHGRI/NIH for sharing a table of corresponding authors of GWAS publications, and Caitlin McHugh of the Department of Biostatistics at the University of Washington in Seattle for formatting the preliminary data files and collating information on the NHGRI list of GWAS publications. We also acknowledge Kathleen Tatem for assistance with graphics and figures.

References

1 Ioannidis, J. P., Castaldi, P. & Evangelou, E. A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst* **102**, 846-858, doi:djq173 [pii] 10.1093/jnci/djq173 (2010).

2 Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. & Contopoulos-Ioannidis, D. G. Replication validity of genetic association studies. *Nat Genet* **29**, 306-309, doi:10.1038/ng749 ng749 [pii] (2001).

3 Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* **39**, 1181-1186, doi:ng1007-1181 [pii] 10.1038/ng1007-1181 (2007).

4 Goldstein, D. B. Common genetic variation and human traits. *N Engl J Med* **360**, 1696-1698, doi:NEJMp0806284 [pii] 10.1056/NEJMp0806284 (2009).

5 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:nature08494 [pii] 10.1038/nature08494 (2009).

6 Aschard, H., Hancock, D. B., London, S. J. & Kraft, P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered* **70**, 292-300, doi:000323318 [pii] 10.1159/000323318 (2010).

7 Khoury, M. J. *et al.* Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. *Am J Epidemiol* **170**, 269-279, doi:kwp119 [pii] 10.1093/aje/kwp119 (2009).

8 Hinch, A. G. *et al.* The landscape of recombination in African Americans. *Nature* **476**, 170-175, doi:nature10336 [pii] 10.1038/nature10336 (2011).

9 Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet* **25**, 489-494, doi:S0168-9525(09)00185-1 [pii] 10.1016/j.tig.2009.09.012 (2009).

10 Bustamante, C. D., Burchard, E. G. & De la Vega, F. M. Genomics for the world. *Nature* **475**, 163-165, doi:475163a [pii] 10.1038/475163a (2011).

11 Manolio, T. A. & Collins, R. Enhancing the feasibility of large cohort studies. *JAMA* **304**, 2290-2291, doi:304/20/2290 [pii] 10.1001/jama.2010.1686 (2010).

12 Cornelis, M. C. *et al.* The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol* **34**, 364-372, doi:10.1002/gepi.20492 (2010).

13 Bennett, S. N. *et al.* Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. *Genet Epidemiol* **35**, 159-173, doi:10.1002/gepi.20564 (2011).

14 Motsinger-Reif, A. A. *et al.* Genome-wide association studies in pharmacogenomics: successes and lessons. *Pharmacogenet Genomics*, doi:10.1097/FPC.0b013e32833d7b45 (2010).

15 Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* **42**, 570-575, doi:ng.610 [pii] 10.1038/ng.610 (2010).

16 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:nature05911 [pii] 10.1038/nature05911 (2007).

17 Yang, W. *et al.* Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet* **6**, e1000841, doi:10.1371/journal.pgen.1000841 (2010).
18 Kooperberg, C., LeBlanc, M. & Obenchain, V. Risk prediction using genome-wide association studies. *Genet Epidemiol* **34**, 643-652, doi:10.1002/gepi.20509 (2010).
19 McCarty, C. A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* **4**, 13, doi:1755-8794-4-13 [pii]
10.1186/1755-8794-4-13 (2011).
20 Yu, W. *et al.* GWAS Integrator: a bioinformatics tool to explore human genetic associations reported in published genome-wide association studies. *Eur J Hum Genet*, doi:ejhg201191 [pii]
10.1038/ejhg.2011.91 (2011).
21 Pendergrass, S. A. *et al.* The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol* **35**, 410-422, doi:10.1002/gepi.20589 (2011).
22 Stover, P. J., Harlan, W. R., Hammond, J. A., Hendershot, T. & Hamilton, C. M. PhenX: a toolkit for interdisciplinary genetics research. *Curr Opin Lipidol* **21**, 136-140, doi:10.1097/MOL.0b013e3283377395 (2010).
23 Moore, G. E. Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). *P Ieee* **86**, 82-85 (1998).
24 Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367-379, doi:nrg2775 [pii]
10.1038/nrg2775 (2010).
25 Yoo, J. & Yoo, H. J. Emerging low energy Wearable Body Sensor Networks using patch sensors for continuous healthcare applications. *Conf Proc IEEE Eng Med Biol Soc* **2010**, 6381-6384, doi:10.1109/IEMBS.2010.5627299 (2010).
26 McGuire, A. L. *et al.* Ethical and practical challenges of sharing data from genome-wide association studies: The eMERGE Consortium experience. *Genome Res* **21**, 1001-1007, doi:gr.120329.111 [pii]
10.1101/gr.120329.111 (2011).
27 Sellers, T. A., Caporaso, N., Lapidus, S., Petersen, G. M. & Trent, J. Opportunities and barriers in the age of team science: strategies for success. *Cancer Causes Control* **17**, 229-237, doi:10.1007/s10552-005-0546-5 (2006).

GWAS Investigator Survey

Dear GWAS Investigator,

The GENEVA and GARNET GWAS consortia and NHGRI have made a commitment to share data and encourage cross-study analyses. We hope that collaboration may help scientists identify genetic loci associated with many different phenotypes and increase the power to study less common outcomes.

We are co-chairs of the GENEVA Phenotype Harmonization committee, and we invite you to participate in this survey of GWAS investigators. This survey asks for general information about your GWAS, the phenotype data you have collected, and your participation in consortia (collaborations of two or more genome-wide association studies [GWAS]).

The purpose of this survey is five-fold:
1. to develop a catalog of GWAS studies and information on the availability of phenotype data;
2. to identify as many consortia as possible so we can develop a catalog of GWAS consortia;
3. to make this information available publicly so it can serve as a resource for GWAS investigators planning research studies or desiring collaboration;
4. to promote future cross-study and cross-consortia collaborations and investigations; and
5. to inform potential funders of the needs of GWAS consortia.

Note that we are also sending out a companion survey to consortia leaders and contacts, asking for specific information on each consortium, its history, membership and primary topic of interest. Should you receive that survey, we hope you will complete it as well.

We would greatly appreciate it if you could take 5 minutes to complete this survey.

Thank you for your interest and support!

Siiri Bennett, MD
University of Washington
Seattle, Washington

Neil Caporaso, MD
National Cancer Institute
Bethesda, Maryland

Question

Your name:

Question

Your email address:

Question

Your Institution:

The NHGRI database indicates that you have contributed to one or more GWAS. If you have only contributed to one GWAS, please indicate that below and answer the rest of the questions for that GWAS. If you have more than one GWAS, please indicate below which GWAS you will be describing below. You are also welcome to repeat this survey for each GWAS.

Question

How many GWAS studies do you have?
Enter an integer (without commas).

Question

Is the GWAS you are describing below:

- Your only GWAS
- Your newest or most recent GWAS
- Your GWAS with the greatest number of citations
- Your GWAS most involved in consortia and collaborations
- Other (please specify):

Question

Name of your GWAS study:

Question

Your study's web address, if it has one:

Question

Primary funding source:

- ☐ NIH grant
- ☐ Government (Intramural) grant
- ☐ Industry grant
- ☐ Other (please specify): _____

Question

Type of GWAS study (tick all that apply):

- ☐ Clinical trial
- ☐ Case-control study
- ☐ Cohort study
- ☐ Family study
- ☐ Other (please specify): _____

Question

Year study started:

- ○ Before 1970
- ○ 1970-1974
- ○ 1975-1979
- ○ 1980-1984
- ○ 1985-1989
- ○ 1990-1994
- ○ 1995-1999
- ○ 2000-2004
- ○ 2005-2009
- ○ 2010-present

Question

Your primary phenotype of interest: (Please tick one)

- ○ Addiction
- ○ Alcohol use and dependence
- ○ Anthropometrics (e.g., height, weight)
- ○ Behavior

- ○ Birth defects
- ○ Cancer
- ○ Cancer treatments
- ○ Cardiovascular conditions (e.g., arrhythmia, blood pressure, peripheral arterial disease)
- ○ Demographics (e.g., age, sex, race/ethnicity)
- ○ Diabetes type 1 or 2
- ○ Diabetic complications
- ○ Environmental exposures (chemical)
- ○ Family history
- ○ Gastrointestinal (GI) disorders or symptoms
- ○ Gynecologic/obstetric conditions
- ○ Imaging findings (e.g., on MRI, CT, X-ray, US, PET scans)
- ○ Infectious diseases (ID) and immunity
- ○ Laboratory findings (e.g., WBC, c-reactive protein, liver function tests)
- ○ Mental health and cognition
- ○ Neurological conditions (e.g., Parkinson's disease, Alzheimer's, stroke)
- ○ Nutrition, diet and dietary supplements
- ○ Occupation/occupational history
- ○ Ocular (eye) conditions
- ○ Oral and dental health
- ○ Pediatric conditions
- ○ Physical activity and exercise
- ○ Preterm birth
- ○ Psychiatric conditions (e.g., ADHD, PTSD, anxiety, depression)
- ○ Psychosocial conditions (e.g., stress, emotional state, quality of life)
- ○ Reproductive health
- ○ Respiratory symptoms/diseases and pulmonary function
- ○ Skin, bone, muscle and joint conditions
- ○ Speech and hearing
- ○ Substance use (other than alcohol and tobacco) and dependence
- ○ Tobacco use
- ○ Other (please specify): _____

Question

Phenotype data collected as part of your GWAS: (Tick all that apply)

- ☐ Addiction
- ☐ Alcohol use and dependence

- ☐ Anthropometrics (e.g., height, weight)
- ☐ Behavior
- ☐ Birth defects
- ☐ Cancer
- ☐ Cancer treatments
- ☐ Cardiovascular conditions (e.g., arrhythmia, blood pressure, peripheral arterial disease)
- ☐ Demographics (e.g., age, sex, race/ethnicity)
- ☐ Diabetes type 1 or 2
- ☐ Diabetic complications
- ☐ Environmental exposures (chemical)
- ☐ Family history
- ☐ Gastrointestinal (GI) disorders or symptoms
- ☐ Gynecologic/obstetric conditions
- ☐ Imaging findings (e.g., on MRI, CT, X-ray, US, PET scans)
- ☐ Infectious diseases (ID) and immunity
- ☐ Laboratory findings (e.g., WBC, c-reactive protein, liver function tests)
- ☐ Mental health and cognition
- ☐ Neurological conditions (e.g., Parkinson's disease, Alzheimer's, stroke)
- ☐ Nutrition, diet and dietary supplements
- ☐ Occupation/occupational history
- ☐ Ocular (eye) conditions
- ☐ Oral and dental health
- ☐ Pediatric conditions
- ☐ Physical activity and exercise
- ☐ Preterm birth
- ☐ Psychiatric conditions (e.g., ADHD, PTSD, anxiety, depression)
- ☐ Psychosocial conditions (e.g., stress, emotional state, quality of life)
- ☐ Reproductive health
- ☐ Respiratory symptoms/diseases and pulmonary function
- ☐ Skin, bone, muscle and joint conditions
- ☐ Speech and hearing
- ☐ Substance use (other than alcohol and tobacco) and dependence
- ☐ Tobacco use
- ☐ Other (please specify): _____

Question

Please indicate below those consortia or collaborations to which you contribute data for cross-study analyses: (Tick all that apply)

- ☐ ABC (African-American Breast Cancer Consortium)
- ☐ ADGC (Alzheimer's Disease Genetic Consortium)
- ☐ ANZgene (Australia and New Zealand Multiple Sclerosis Genetics Consortium)
- ☐ Asian Barrett's Consortium
- ☐ Asian Cohort Consortium
- ☐ BC2OS (Breast Cancer Consortium for Outcomes And Survival)
- ☐ BCAC (Breast Cancer Association Consortium)
- ☐ B-CFR (Breast Cancer Family Registry)
- ☐ BEACON (International Barrett's and Esophageal Adenocarcinoma Consortium)
- ☐ Body Mass Index (BMI) and All Cause Mortality Pooling Project
- ☐ BPC3 (Breast and Prostate Cancer and Hormone-Related Gene Variant Study)
- ☐ BPC3 (Breast and Prostate Cancer Cohort Consortium)
- ☐ BTEC (Brain Tumor Epidemiology Consortium)
- ☐ CADISP (Cervical Artery Dissections and Ischemic Stroke Patients)
- ☐ CALiCo Consortium (Genetic Epidemiology of Causal Variants Across the Life Course)
- ☐ CaP Genes (Prostate Cancer Genetics Study)
- ☐ CARe (Candidate-gene Association REsource)
- ☐ C-CFR (Colon Cancer Family Registry)
- ☐ CGASP (Consortium of Genetic Association of Smoking Related Phenotypes)
- ☐ CGN (Cancer Genetics Network)
- ☐ CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology)
- ☐ CIMBA (Consortium of Investigators of Modifiers of BRCA1/2)
- ☐ CKDGen Consortium
- ☐ CLIC (Childhood Leukemia International Consortium)
- ☐ COGENT (COlorectal cancer GENeTics)
- ☐ Cohort Consortium
- ☐ CRC (Chronic Lymphocytic Leukemia Research Consortium)
- ☐ DentalSCORE (Dental Strategies Concentrating on Risk Evaluation)
- ☐ DGI (Diabetes Genetics Initiative)
- ☐ DIAGRAM (Diabetes Genetics Replication And Meta-analysis Consortium)
- ☐ E2C2 (Epidemiology of Endometrial Cancer Consortium)
- ☐ EADI (European Alzheimer's Disease Initiative)
- ☐ eMERGE (Electronic Medical Records & Genomics)
- ☐ ENGAGE (European Network of Genomic and Genetic Epidemiology)
- ☐ ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles Consortium)
- ☐ EUROCRAN (European Collaboration on Craniofacial Anomalies)

- ☐ GAPPS (Global Alliance to Prevent Prematurity and Stillbirth)
- ☐ GARNET (Genomics and Randomized Trials Network, or Genome-wide Association Research Network into Effects of Treatment)
- ☐ GECCO (Colorectal Cancer GWAS Consortium)
- ☐ GEFOS (Genetic Factors of Osteoporosis Consortium)
- ☐ GELCC (Genetic Epidemiology of Lung Cancer)
- ☐ GEM (Genes, Environment and Melanoma)
- ☐ GEMINI (Gene-Environment Meta-analysis of NIcotine dependence) Consortium
- ☐ GENEVA (GENe EnVironment Association studies)
- ☐ GenoMEL (Melanoma Genetics Consortium)
- ☐ GIANT (Genome-wide Investigation of ANThropometric measures)
- ☐ GLIOGENE
- ☐ Global BPGen Consortium
- ☐ Global Lipid Genetics Consortium
- ☐ I4C (International Childhood Cancer Cohort Consortium)
- ☐ ICPCG (International Consortium on Prostate Cancer Genetics)
- ☐ IGAP (International Genomics of Alzheimer's Project)
- ☐ IIBDGC (International Inflammatory Bowel Disease Genetics Consortium)
- ☐ ILCCO (International Lung Cancer Consortium)
- ☐ IMMC (International Multiple Myeloma Consortium)
- ☐ INHANCE (International Head and Neck Cancer Consortium)
- ☐ InterLymph (International Consortium on Lymphoma Epidemiologic Studies)
- ☐ International Cancer Genome Consortium
- ☐ International Headache Genetics Consortium
- ☐ International Kawasaki Disease Genetics Consortium
- ☐ International Type 2 Diabetes Consortium
- ☐ ISGC (International Stroke Genetics Consortium)
- ☐ LACE (Latin American Cancer Epidemiology Consortium)
- ☐ MADCaP (Men of African Descent and Prostate Cancer)
- ☐ MAGIC (The Meta-Analyses of Glucose and Insulin-related traits Consortium)
- ☐ MECC (Molecular Epidemiology of Colorectal Cancer)
- ☐ NEIGHBOR (National Eye Institute Glaucoma Human Genetics CollaBORation)
- ☐ NGFN (German National Genome Research Network)
- ☐ OCAC (Ovarian Cancer Association Consortium)
- ☐ Ovarian Cancer Cohort Consortium
- ☐ P3G Consortium (Public Population Project in Genomics)
- ☐ PACGENE (Pancreatic Cancer Genetic Epidemiology Consortium)
- ☐ PAGE (Population Architecture using Genomics and Epidemiology)
- ☐ PANC4 (Pancreatic Cancer Case Control Consortium)

- ☐ Pancreatic Cancer Cohort Consortium
- ☐ PBTC (Pediatric Brain Tumor Consortium)
- ☐ POCRC (Pacific Ovarian Cancer Research Consortium)
- ☐ Post-Genome Wide Association Initiative
- ☐ PREGENIA (Preterm Birth and Genetics International Alliances)
- ☐ PROSE (Prevention and Observation of Surgical Endpoints)
- ☐ Radiogenomics Consortium
- ☐ ReproGen Consortium
- ☐ SHARe (SNP Health Association Research)
- ☐ SHARP (SHARe Asthma Resource Project)
- ☐ SpiroMeta Consortium
- ☐ STAMPEED (SNP Typing for Association with Multiple Phenotypes from Existing Epidemiologic Data)
- ☐ STOMP (Study of Tobacco in Minority Populations) Genetics Consortium
- ☐ SUNLIGHT Consortium (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits)
- ☐ TAG (The Tobacco, Alcohol and Genetics Consortium)
- ☐ Type 1 Diabetes Genetics Consortium
- ☐ Urinary Bladder Cancer Consortium
- ☐ VDPP (Vitamin D Pooling Project)
- ☐ WECARE (Women, Cancer and Radiation Exposure)
- ☐ WTCCC (Wellcome Trust Case-Control Consortium)
- ☐ TRICL (Transdisciplinary Research in Cancer of the Lung)
- ☐ Other (please specify): [ ]

Please list below any consortia or collaborations not listed above that you feel should be added to the list:

Question

1. Consortium name:

[ ]

Question

Consortium web address:

[ ]

Question

Contact person:

Question

Contact person email address:

Question

2. Consortium name:

Question

Consortium web address:

Question

Contact person email address:

Question

Contact person:

Question

3. Consortium/collaboration name:

Question

Consortium web address:

Question

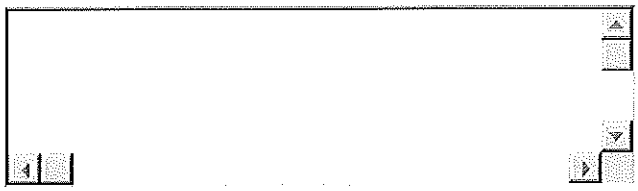
Contact person:

Question

Contact person email address:

Question

We welcome your comments and suggestions. Please feel free to comment below.



Thank you for your time and support! We will inform you by email when the results of the survey are available.

GWAS Consortium Survey

Dear GWAS Investigator,

The GENEVA and GARNET GWAS consortia and NHGRI have made a commitment to share data and encourage cross-study analyses. We hope that collaboration may help scientists identify genetic loci associated with many different phenotypes and increase the power to study less common outcomes.

We are co-chairs of the GENEVA Phenotype Harmonization committee. According to our information, you are a leader or contact person for one or more GWAS consortia (collaborations of two or more genome-wide association studies [GWAS]), and we invite you to participate in this survey of GWAS consortia. This survey asks for general information about your consortium, its history, membership, and primary topic of interest.

The purpose of this survey is five-fold:
1. to develop a catalog of GWAS consortia with information on each;
2. to assess the growth, number, size and characteristics of consortia;
3. to make this information available publicly so it can serve as a resource for GWAS investigators planning research studies or desiring collaboration;
4. to promote future cross-study and cross-consortia collaborations and investigations; and
5. to inform potential funders of the needs of GWAS consortia.

Note that we are also sending out a companion survey to all GWAS publication corresponding authors, asking for specific information on their GWAS, the phenotype data collected, and their participation in consortia. Should you receive that survey, we hope you will complete it as well.

We would greatly appreciate it if you could take 5 minutes to complete this survey.

Thank you for your interest and support!

Siiri Bennett, MD
University of Washington
Seattle, Washington

Neil Caporaso, MD
National Cancer Institute
Bethesda, Maryland

Question

Your name:

Question

Your email address:

Question

Your institution:

Question

What is the name of your consortium? (If you participate in more than one consortium, please consider completing the survey more than once, once for each consortium. Thank you!)

Question

What is your consortium's website URL? (If your consortium does not have a website, please enter NA)

Question

What is the month and year the consortium formed? (MM/YYYY)

Question

What is the status of the consortium?

- It still exists as a consortium
- The consortium no longer exists

Question

What was the month and year the consortium closed? (MM/YYYY)

Question

What are the diseases or conditions around which the consortium is structured? (Please list)

Question

How many GWAS studies contribute to or participate in the consortium?
Enter an integer (without commas).

Question

Approximately how many subjects total are represented by the consortium? (This is the sum of the number of subjects in each study contributing to the consortium.)
Enter an integer (without commas).

Question

On what basis does the consortium include participating studies? (Tick all that apply)

- ☐ Participating studies' primary phenotype is the consortium's phenotype of interest
- ☐ Participating studies' primary phenotype is not the consortium's phenotype of interest, but the studies have data relevant to the aim of the consortium
- ☐ Participating studies use data obtained from another source, such as dbGaP
- ☐ Other (please specify):

Question

What is the primary funding source for the consortium? (This question refers to the funding for the consortium, not the individual participating GWAS. Tick all that apply.)

- ☐ NIH grant
- ☐ Government (Intramural) grant
- ☐ Industry grant
- ☐ None
- ☐ Other (please specify):

Question

How are the activities of the consortium coordinated and managed? (Tick all that apply)

- ☐ Through an independent Coordinating Center
- ☐ With support from a specific institution
- ☐ Through periodic in-person meetings
- ☐ Through conference calls
- ☐ Through the consortium's website
- ☐ Use of other electronic media (social media, etc.)
- ☐ Ad hoc organization as required
- ☐ Other (please specify): _____

Question

What genotyping platforms are used by the GWAS studies in your consortium? (Tick all that apply)

- ☐ Affymetrix
- ☐ Illumina
- ☐ Perlegen
- ☐ Other (please specify): _____

Question

What types of biospecimens have been used for genotyping in the consortium? (Tick all that apply)

- ☐ Blood-based DNA
- ☐ Buccal-based DNA
- ☐ Tissue-based DNA
- ☐ None (analyses are based on metadata or in silico)
- ☐ Don't know
- ☐ Other (please specify): _____

Question

Please indicate below if biospecimens are available on some or all of the individuals in the consortium. (Tick all that apply)

- ☐ Blood
- ☐ Urine
- ☐ Buccal
- ☐ Tissue
- ☐ Don't know
- ☐ Other (please specify):

Question

What methods are used for cross-study analyses? (Tick all that apply)

- ☐ Pooled analysis (analysis of pooled individual-level data)
- ☐ Meta-analysis (analysis of summary statistics)
- ☐ Don't know
- ☐ Other (please specify):

Question

What was the month and year of the first publication deriving from this consortium? (MM/YYYY)

Question

Approximately how many publications have come out of the consortium (from the collaboration of participating studies, not from individual participating studies)?
Enter an integer (without commas).

The following six questions ask about the population on which the consortium (and its participating GWAS) is focused:

Question

1. What is the gender of the consortium's population?

- ○ Men only
- ○ Women only
- ○ Both men and women

Does this include pregnant women?

- ○ Yes
- ○ No
- ○ Don't know

Question

2. What is the age of the consortium's population? (Tick all that apply)

- ☐ Infant
- ☐ Pediatric
- ☐ Teen
- ☐ Young adult
- ☐ Middle-aged
- ☐ Elderly
- ☐ Other age range (please specify): _____

Question

3. The consortium's subjects are of which race? (Tick all that apply)

- ☐ White/ Caucasian ancestry
- ☐ Asian ancestry
- ☐ Black/ African ancestry
- ☐ Pacific Islander/ Native Hawaiian
- ☐ American Indian or Alaskan Native
- ☐ Biracial or multiracial
- ☐ Unknown or missing race
- ☐ Other (please specify): _____

Question

4. The consortium's subjects are of which ethnicity? (Tick all that apply)

- ☐ Self-identified Hispanic or Latino ethnicity
- ☐ Not of Hispanic or Latino ethnicity
- ☐ Ethnicity not specified or unknown

Question

5. Do the GWAS included in your consortium focus on any of the following populations? (Tick all that apply)

- ☐ Twins
- ☐ Related individuals
- ☐ Families (includes individuals from more than one generation, and can include both related and unrelated individuals)
- ☐ No focus on related individuals or special populations
- ☐ Other special or restricted population (please specify): _____

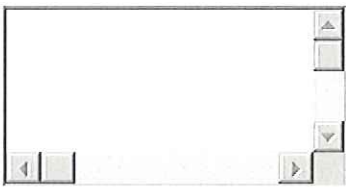
Question

6. Geographic area covered by the consortium: (Tick all that apply)

- ☐ North America
- ☐ Central American & the Caribbean
- ☐ South America
- ☐ Europe
- ☐ Africa
- ☐ Eastern Asia (e.g., China, Japan, Korea)
- ☐ Southern Asia (e.g., India, Sri Lanka, Pakistan, Bangladesh)
- ☐ Southeast Asia (e.g., Vietnam, Thailand, Singapore, Malaysia)
- ☐ Western Asia (e.g., Turkey, Israel, United Arab Emirates)
- ☐ Oceania (e.g., Melanesia, Polynesia)
- ☐ Australia/New Zealand
- ☐ Don't know
- ☐ Other (please specify): _____

Question

We welcome your comments and suggestions. Please feel free to comment below.



Thank you for your time and support! We will inform you by email when the results of the survey are available.

Questions or Comments?
Contact Siiri Bennett at siirib@u.washington.edu

Table 1. NHGRI GWAS publications with the largest sample sizes

First Author	Date	Journal	Link	Publication title	Disease Trait	Initial Sample Size	Replication Sample Size	Total Sample Size	Race	Subject source (consortium/ study name)	No. of studies*
Speliotes	10/10/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20935630	Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index	Body mass index	Up to 123,865 European ancestry individuals	Up to 125,931 European ancestry individuals	249,796	Caucasian	GIANT	46/42
Heid	10/10/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20935629	Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution	Waist-hip ratio	Up to 77,167 European descent individuals	Up to 113,636 European descent individuals	190,803	Caucasian	GIANT	32/29
Allen	9/29/2010	Nature	http://www.ncbi.nlm.nih.gov/pubmed/20881960	Hundreds of variants clustered in genomic loci and biological pathways affect human height	Height	133,653 Caucasian European ancestry individuals	50,074 Caucasian European ancestry individuals	183,727	Caucasian	GIANT	46/15
Liu	4/25/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20418889	Meta-analysis and imputation refines the association of 15q25 with smoking quantity	Smoking behavior	41,150 European descent individuals	120,516 European descent individuals	161,666	Caucasian	WTCCC included in studies	20
Newton-Cheh	5/10/2009	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/19430483?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_Re	Genome-wide association study identifies eight loci associated with blood pressure.	Diastolic blood pressure	34,433 individuals	Up to 100,347 white individuals, up to 12,889 Indian Asian individuals	147,669	NA	Global BPgen	17/13

			sultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum								
Schunkert	3/6/2011	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/21378990	Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease	Coronary heart disease	22,233 European ancestry cases, 64,762 European ancestry controls	28,834 cases, 27,674 controls, 2,115 individuals.	145,618	Caucasian	CARDIoGRAM consortium	14
The Tobacco and Genetics Consortium	4/25/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20418890	Genome-wide meta-analyses identify multiple loci associated with smoking behavior	Smoking behavior	Up to 74,035 European ancestry individuals	Up to 68,988 participants	143,023	Caucasian	TAG, ENGAGE, Ox-GSK	16
Voight	6/27/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20581827	Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis	Type 2 diabetes	8,130 European descent cases, 38,987 European descent controls	Up to 34,412 European descent cases, 59,925 European descent controls	141,454	Caucasian	WTCCC, DGI, FUSION, and others--DIAGRAM+	8
Lindgren	6/26/2009	PLoS Genet	http://www.ncbi.nlm.nih.gov/pubmed/19557161?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum	Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution.	Adiposity	38,580 European-ancestry individuals	Up to 102,064 European-ancestry individuals	140,644	NA	GIANT, CHARGE	8
Teslovich	8/5/2010	Nature	http://www.ncbi.nlm.nih.gov/pubmed/20686565	Biological, clinical and population relevance of 95 loci for blood lipids	Blood lipid traits	Up to 100,184 European descent individuals	9,705 South Asians, 15,046 East Asians, 8,061 African	140,059	Caucasian	CARDIoGRAM, CARe, deCODE included	46

							Americans, 7,063 Europeans				
Dupuis	1/17/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20081858?itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum&ordinalpos=1	New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.	Fasting glucose-related traits	up to 46,186 European descent individuals	up to 76,558 European ancestry individuals	122,744	Caucasian	MAGIC	21/20
Elks	11/21/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/21102462	Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies	Menarche (age at onset)	87,802 European ancestry women	Up to 14,731 women	102,533	Caucasian	ReproGen	32
Thorleifsson	12/14/2008	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/19079260?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum	Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity.	Body mass index	80,969 individuals	11,036 individuals	92,005	NA	GIANT, DGI included	19
Willer	12/14/2008	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/19079261?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_Defa	Six new loci associated with body mass index highlight a neuronal influence on body weight regulation.	Body mass index	32,387 individuals	59,092 individuals	91,479	NA	GIANT	15/14

			ultReportPanel.PubmedRVDocSum								
Qi	4/23/2010	Hum Mol Genet	http://www.ncbi.nlm.nih.gov/pubmed/20418489	Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes	Type 2 diabetes	2,591 European ancestry cases, 3,052 European ancestry controls	10,870 European origin cases, 73,735 European origin controls	90,248	Caucasian	MAGIC, DIAGRAM	21
Kottgen	4/11/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20383146	New loci associated with kidney function and chronic kidney disease	Chronic kidney disease	Up to 67,093 European ancestry individuals	Up to 22,982 European ancestry individuals	90,075	Caucasian	CKDGen consortium	20
Zeggini	3/30/2008	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/18372903?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum	Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes.	Type 2 diabetes	4,549 cases,5,579 controls	24,194 cases,55,598 controls	89,920	NA	DGI, FUSION, WTCCC	3
Thorgeirsson	4/25/2010	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20418888	Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior	Smoking behavior	31,266 European ancestry individuals	54,731 European descent individuals	85,997	Caucasian	ENGAGE, TAG, Ox-GSK	12
Dehghan	2/7/2011	Circulation	http://www.ncbi.nlm.nih.gov/pubmed/21300955	Meta-Analysis of Genome-Wide Association Studies in >80 000 Subjects Identifies Multiple Loci for C-Reactive Protein Levels	C-reactive protein	66,185 European ancestry individuals	16,540 European ancestry individuals	82,725	Caucasian	CHARGE, EUROSPAN	15/10
Loos	5/4/2008	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/18454148?or	Common variants near MC4R are associated	Body mass index	16,876 individuals	60,352 individuals	77,228	NA	WTCCC included in studies	7

			dinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum	with fat mass, weight and risk of obesity.							
Repapi	12/13/2009	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/20010834?itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum&ordinalpos=1	Genome-wide association study identifies five loci associated with lung function.	Pulmonary function	20,288 European ancestry individuals	up to 54,276 European ancestry individuals	74,564	Caucasian	SpiroMeta, CHARGE	14/4
The Coronary Artery Disease (C4D) Genetics Consortium	3/6/2011	Nat Genet	http://www.ncbi.nlm.nih.gov/pubmed/21378988	A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease	Coronary heart disease	8,424 European ancestry cases, 7,268 European ancestry controls, 6,996 South Asian cases, 7,794 South Asian controls	18,049 European ancestry cases, 16,357 European ancestry controls, 3,359 South Asian cases, 2,828 South Asian controls	71,075	Caucasian	C4D, PROCARDIS	4
Heard-Costa	6/26/2009	PLoS Genet	http://www.ncbi.nlm.nih.gov/pubmed/19557197?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum	NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium.	Waist circumference	31,373 Caucasian individuals	38,641 Caucasian individuals	70,014	Caucasian	CHARGE	8
Gudbjartsson	7/29/2010	PLoS Genet	http://www.ncbi.nlm.nih.g	Association of variants at	Chronic kidney	2,903 Icelandic	300 Icelandic	68,439	Caucasian	<appears to be 2	2

			ov/pubmed/20686651	UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases	disease and serum creatinine concentration	CKD cases, 35,818 Icelandic controls, 22,256 Icelandic subjects with serum creatinine	CKD cases, 2,964 Icelandic controls, 2,379 Icelandic subjects with serum creatinine, 1,819 Dutch subjects with serum creatinine			individual studies>	
Panoutsopoulou	12/21/2010	Ann Rheum Dis	http://www.ncbi.nlm.nih.gov/pubmed/21177295	Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study	Osteoarthritis	3,177 European descent cases, 4,894 European descent controls	10,525 European descent cases, 48,392 European descent controls	66,988	Caucasian	arcOGEN, with WTCCC	5

* If 2 numbers are given, these are the number of studies in 1st and 2nd stage analyses respectively

Table 2. Consortia funding sources

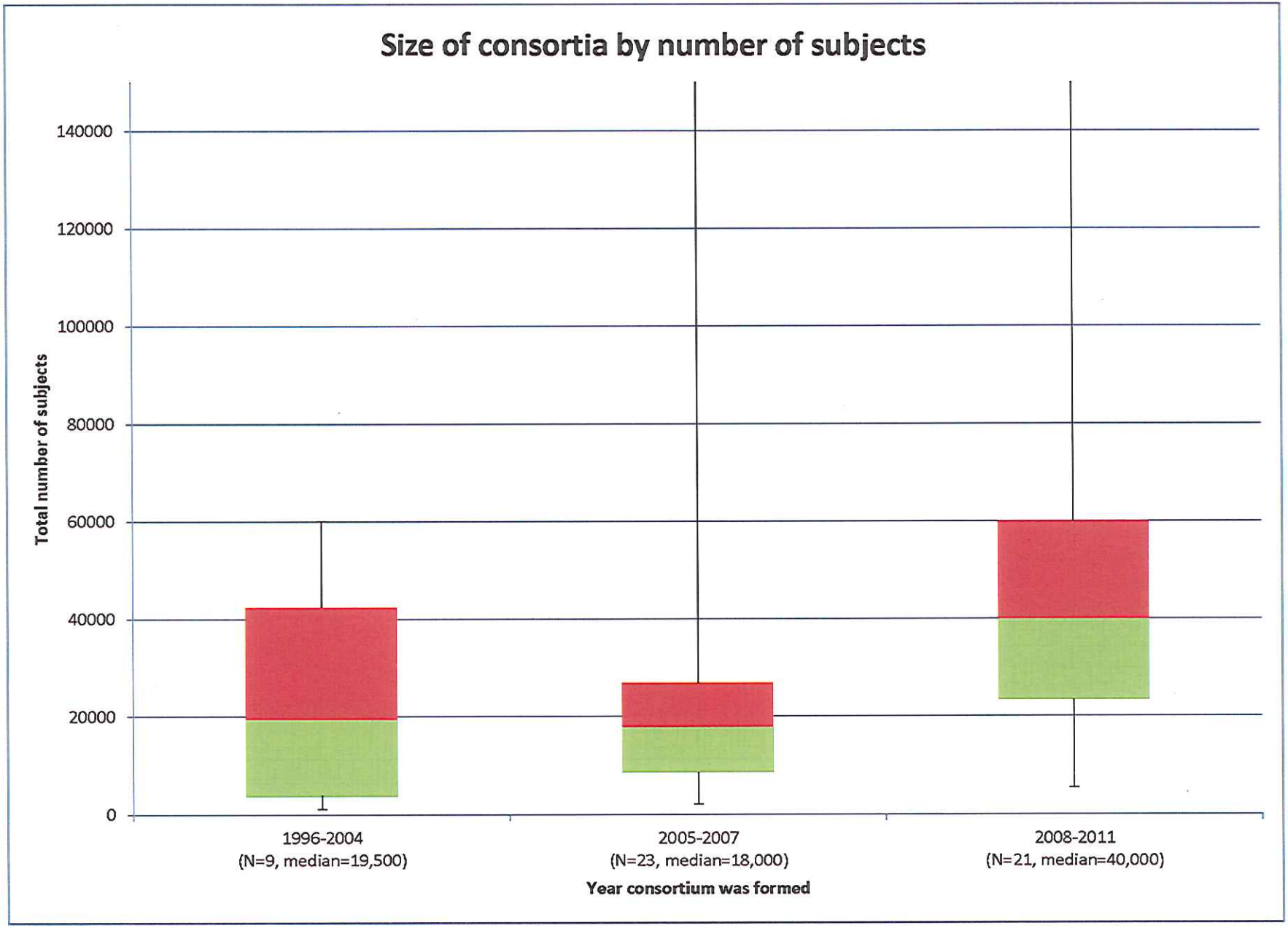
Funding source	Number of consortia (N=51)*
National Institutes of Health (NIH)	33 (65%)
Government (intramural) grant	6 (12%)
Industry	3 (6%)
Other:	
Wellcome Trust	3 (6%)
Foundation	2 (4%)
European funding agency	6 (12%)
Development funds	1 (2%)
Contract	1 (2%)
University funds	1 (2%)
Unidentified funding source	3 (6%)
None identified	10 (20%)

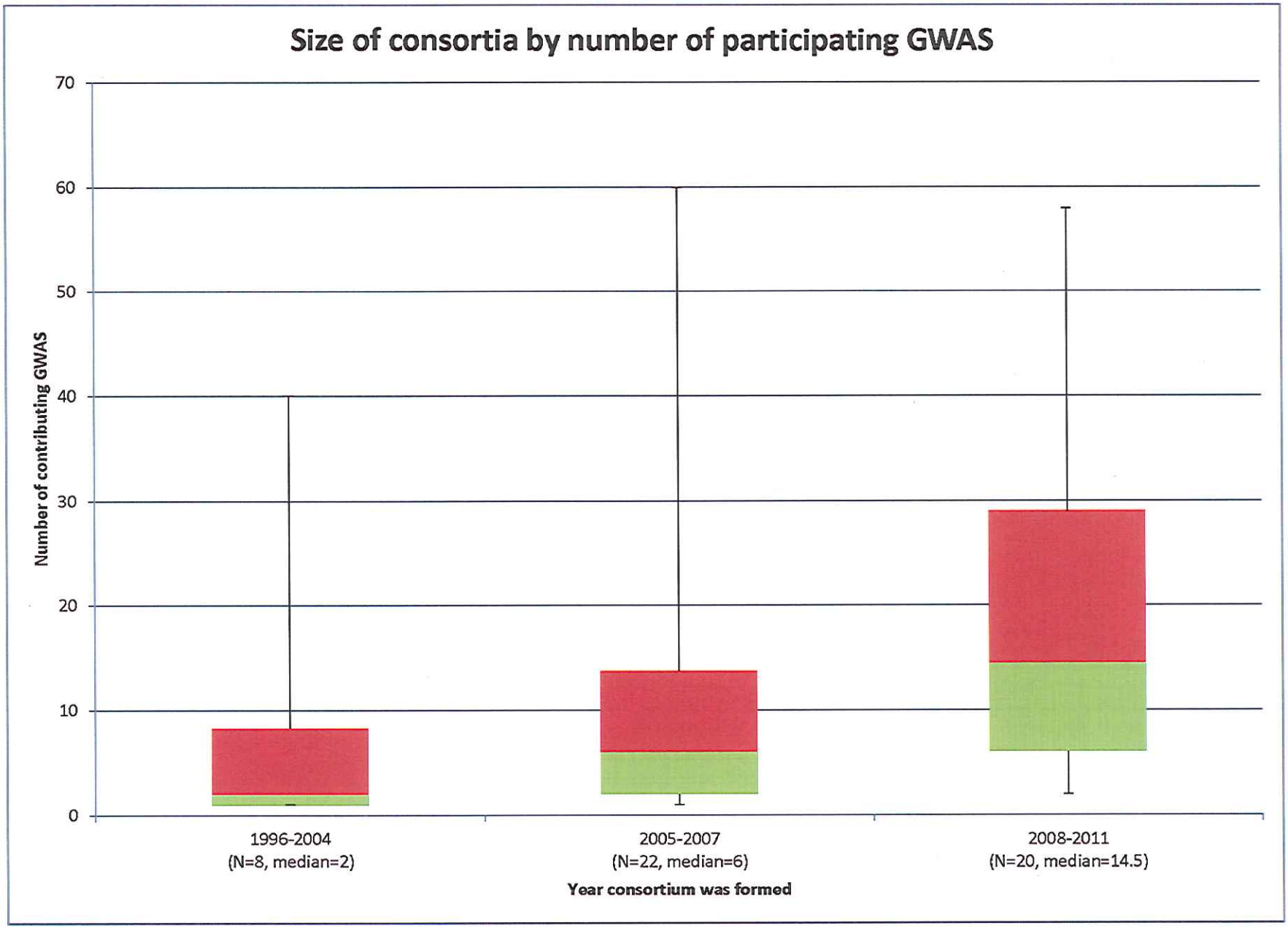
* Consortia could indicate more than one funding source

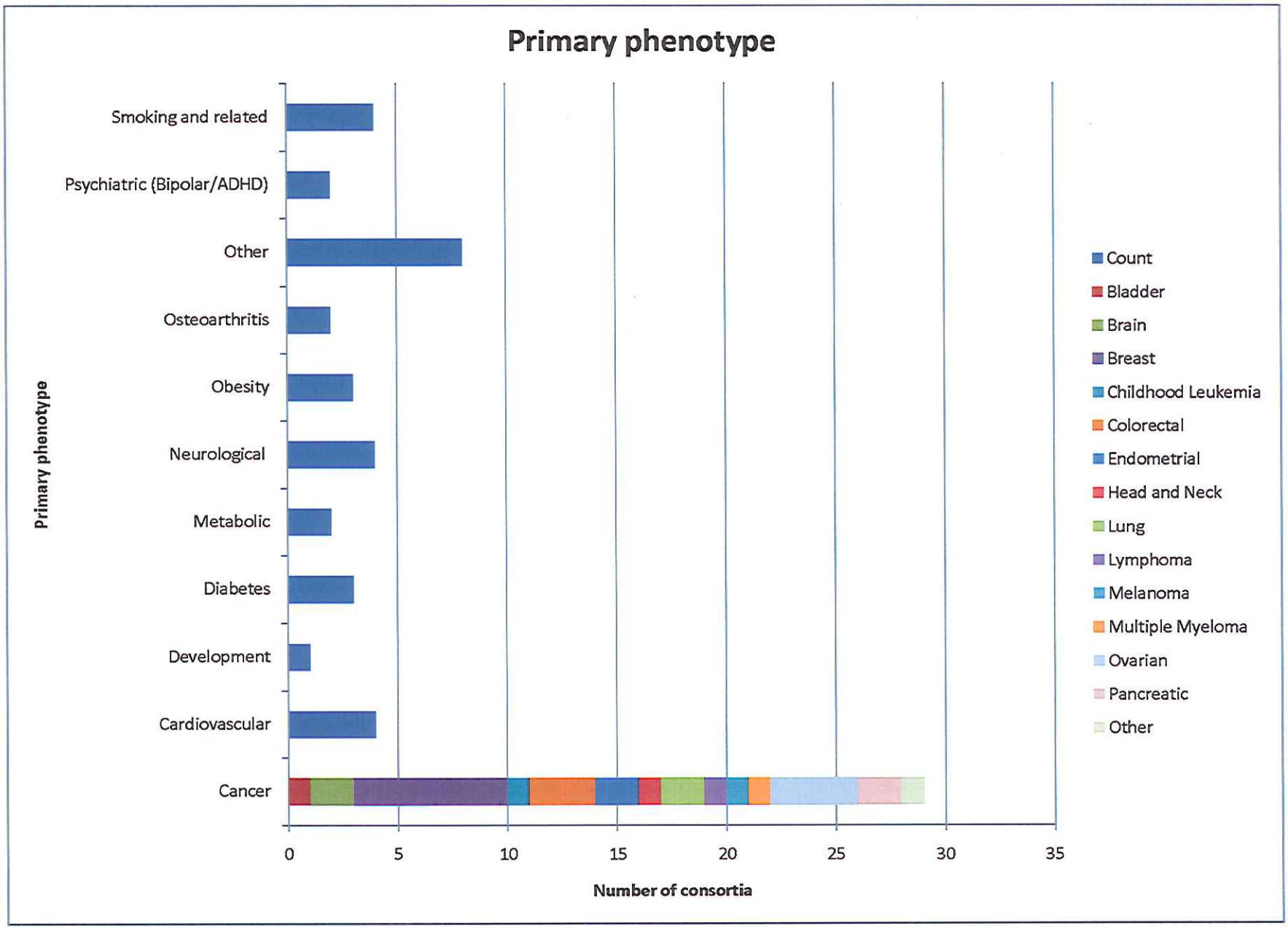
Table 3. Advantages and disadvantages of consortia

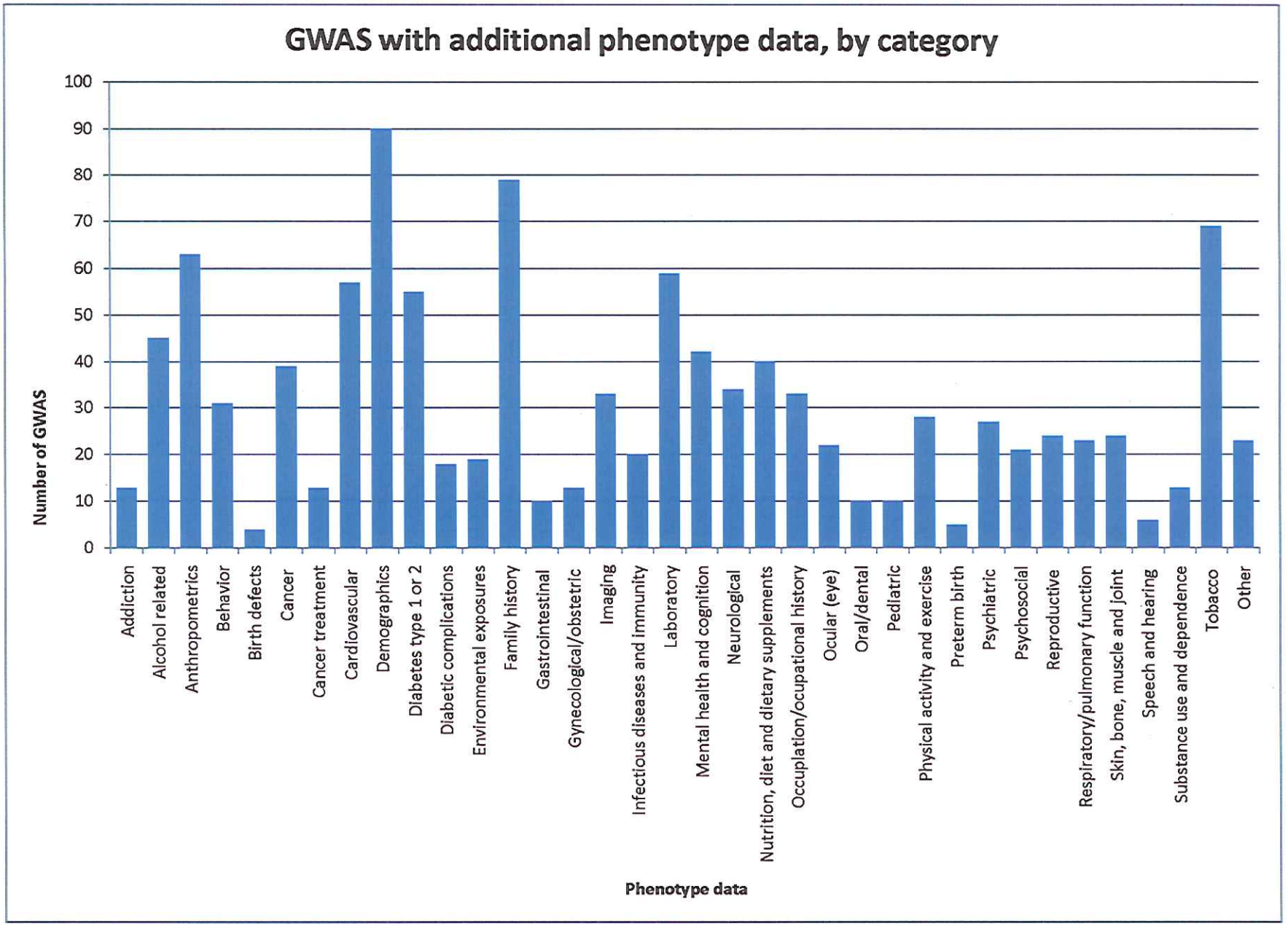
Advantages
Sample size (Burton PR, et al. 2008)
Data harmonization (Bennett et al. 2011)
Interdisciplinary participation (enhanced expertise in diverse areas)
Improved quality control (Laurie, 2010)
Standardized procedures for analyses, imputation, etc.
Standard policies to deal with authorship, IRB, consent, intellectual property issues
Policies favor depositing data in public repositories such as dbGaP
Reduced marginal costs

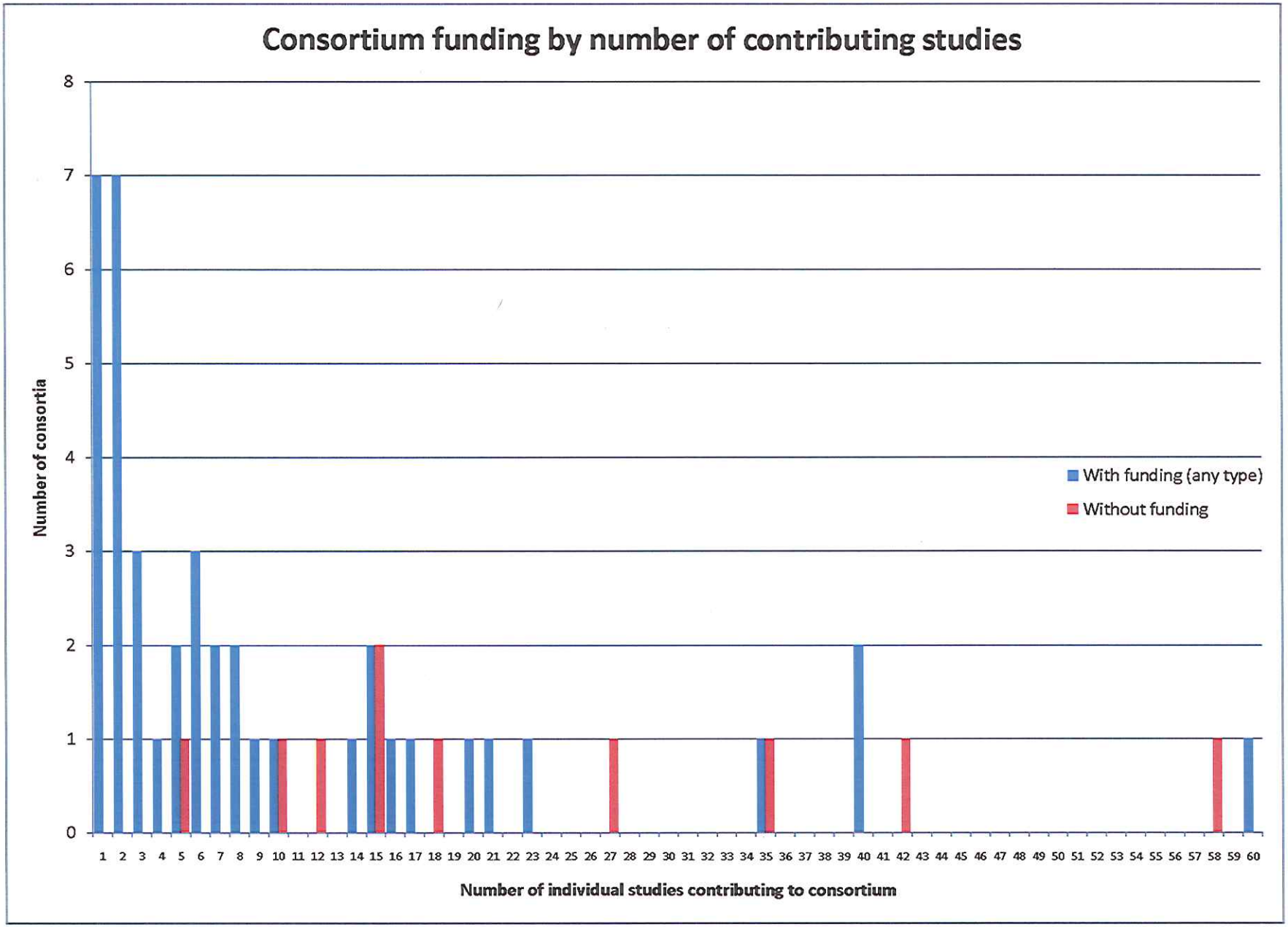
Disadvantages
Diminished autonomy and control
Complex organization, particularly when dealing with consent issues
Ethical issues
Authorship sharing/dilution
Need to enhance participation of younger scientists

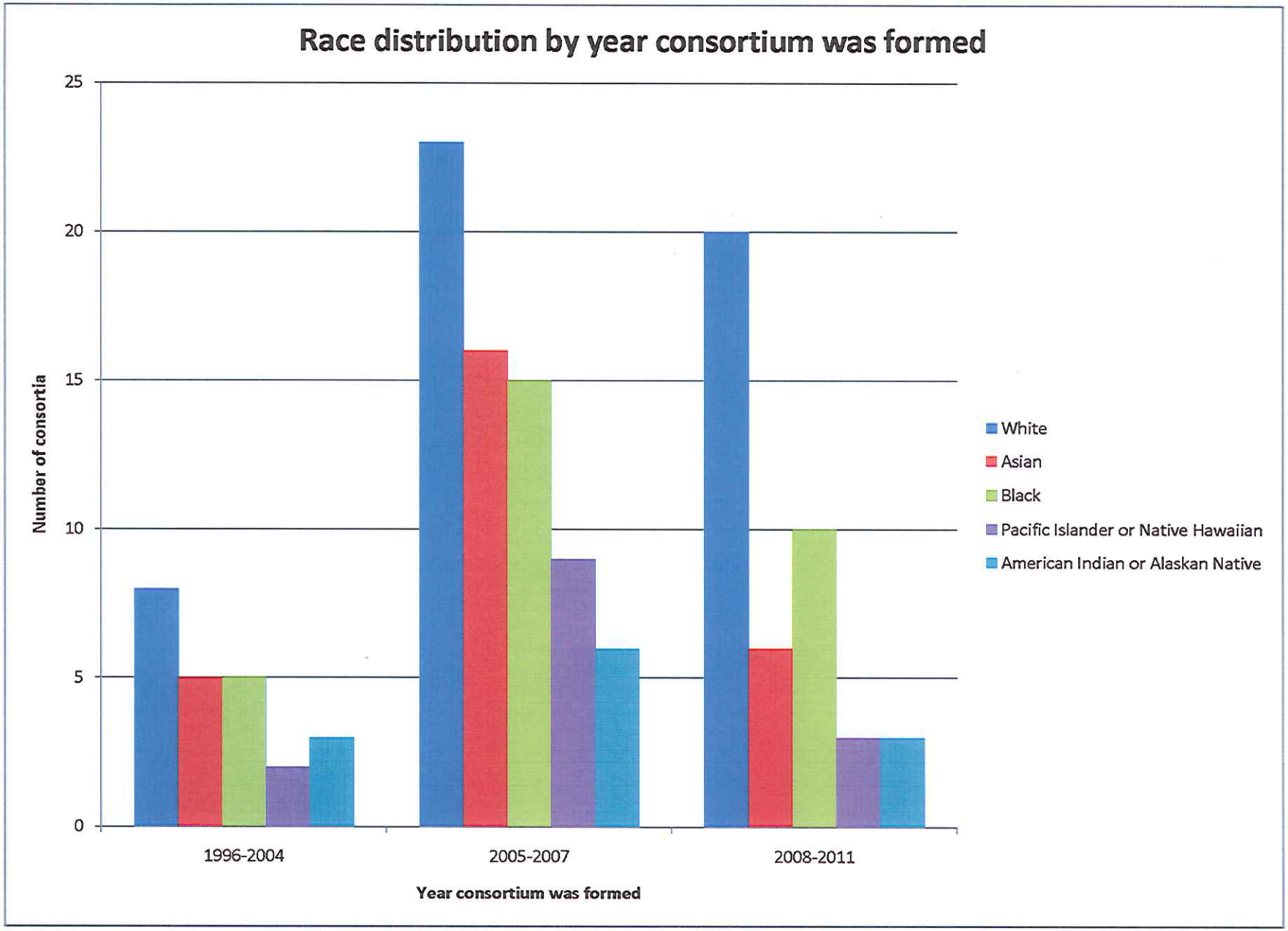

Size of consortia by number of subjects
Total number of subjects
140000
120000
100000
80000
60000
40000
20000
0
1996-2004
(N=9, median=19,500)
2005-2007
(N=23, median=18,000)
2008-2011
(N=21, median=40,000)
Year consortium was formed

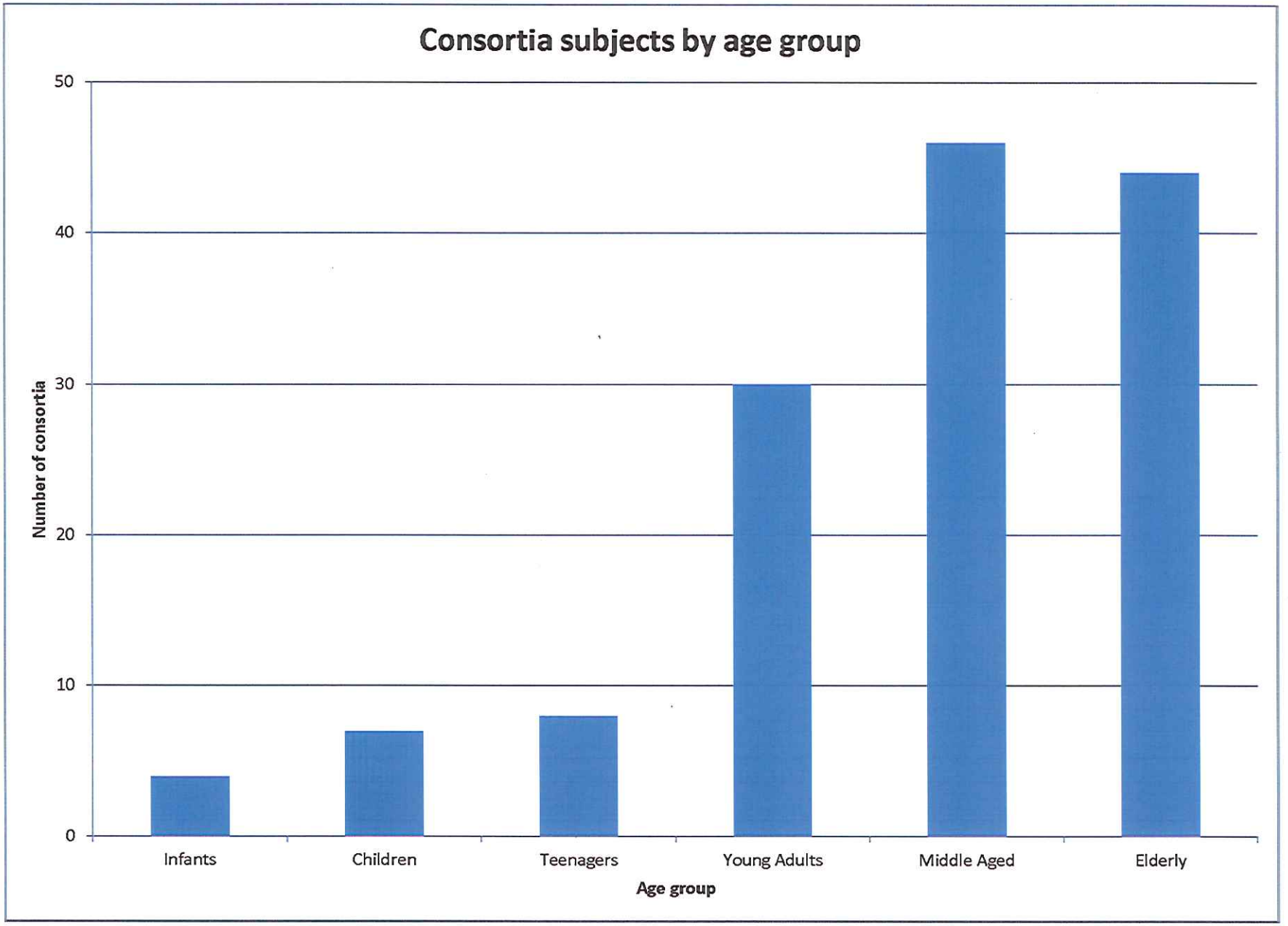

Size of consortia by number of participating GWAS
70
60
50
40
30
20
10
0
Number of contributing GWAS
1996-2004
(N=8, median=2)
2005-2007
(N=22, median=6)
2008-2011
(N=20, median=14.5)
Year consortium was formed

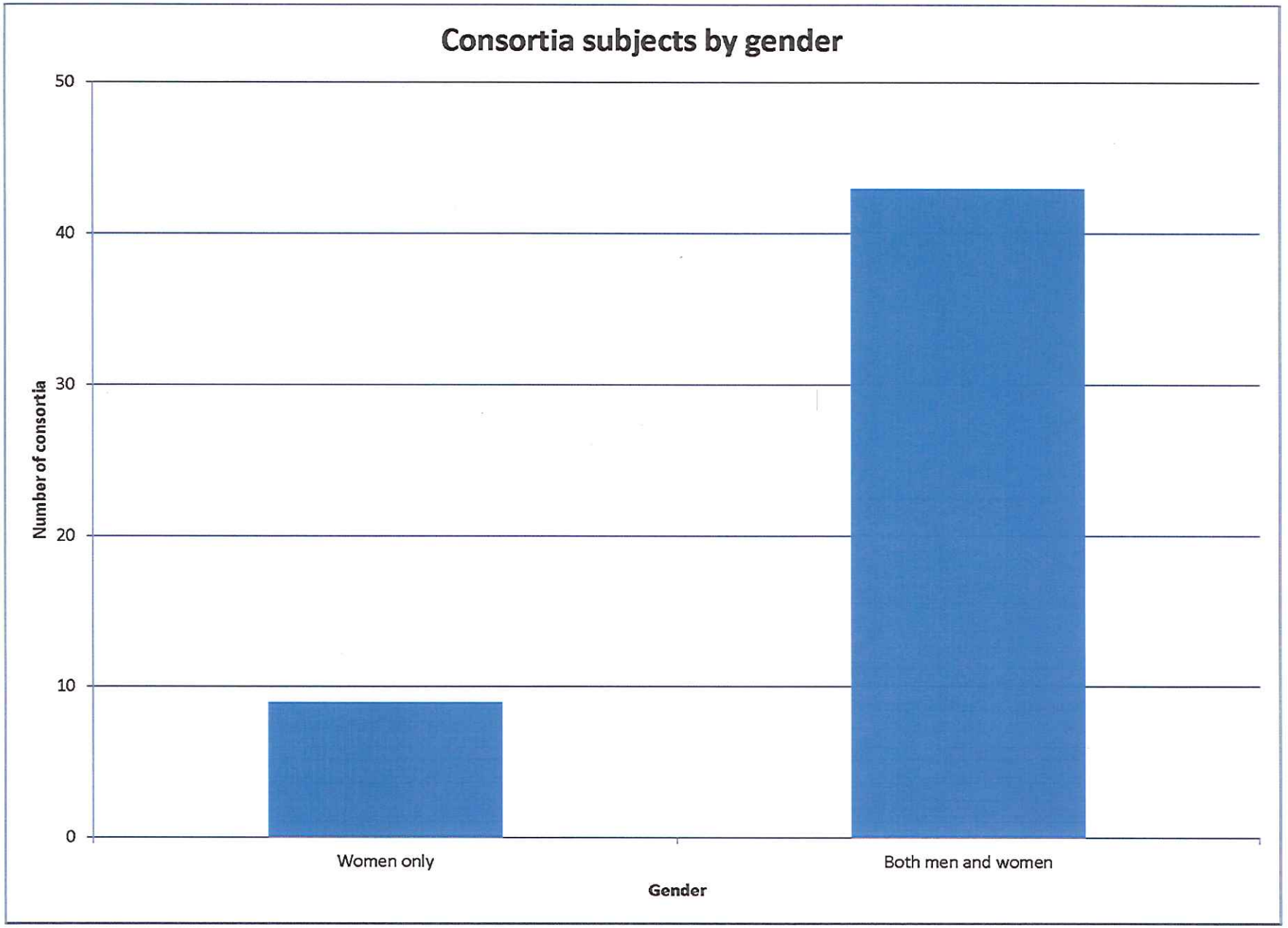

Primary phenotype
Smoking and related
Psychiatric (Bipolar/ADHD)
Other
Osteoarthritis
Obesity
Neurological
Metabolic
Diabetes
Development
Cardiovascular
Cancer
Primary phenotype
0
5
10
15
20
25
30
35
Number of consortia
Count
Bladder
Brain
Breast
Childhood Leukemia
Colorectal
Endometrial
Head and Neck
Lung
Lymphoma
Melanoma
Multiple Myeloma
Ovarian
Pancreatic
Other

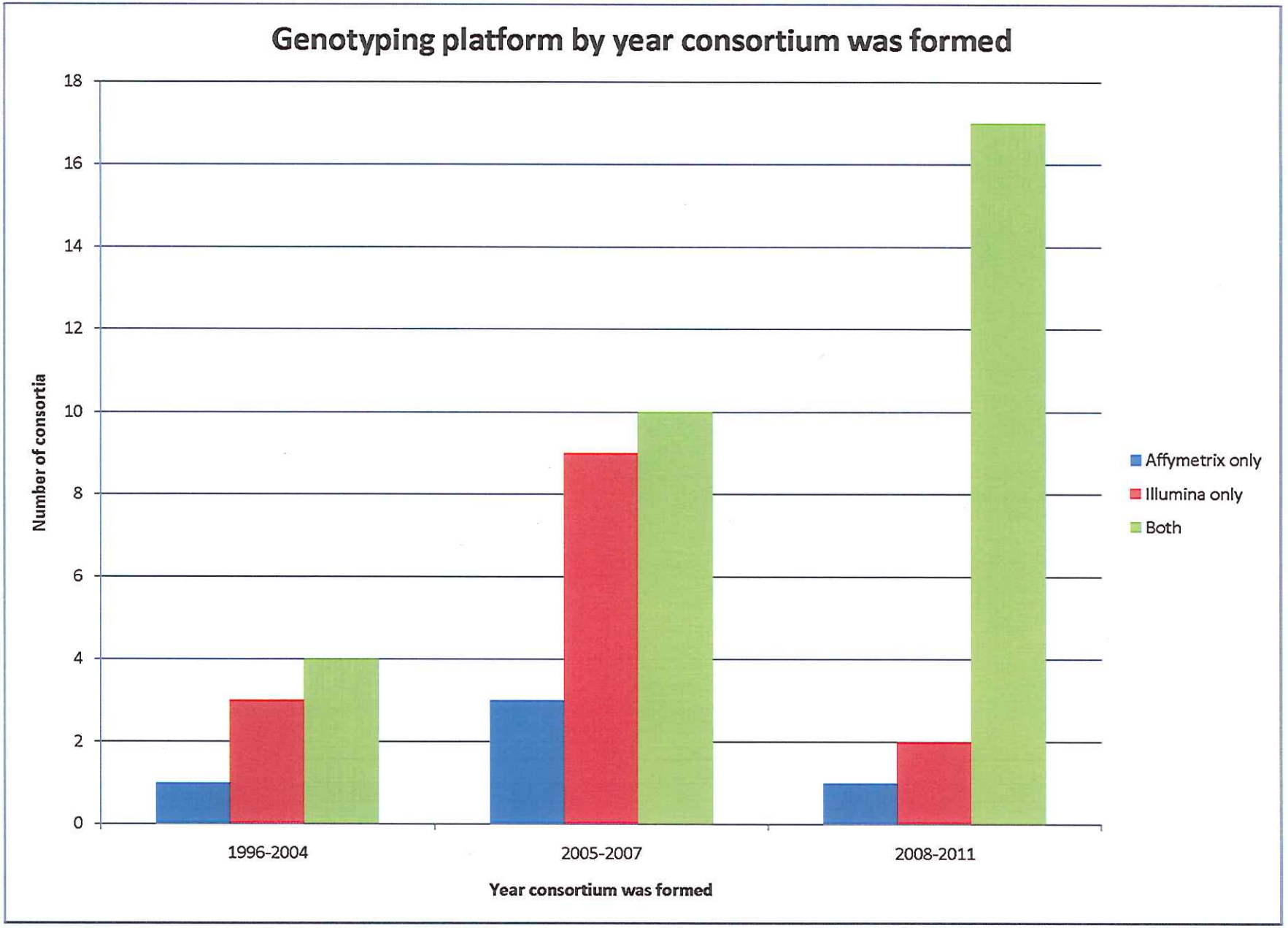

GWAS with additional phenotype data, by category
Number of GWAS
0
10
20
30
40
50
60
70
80
90
100
Addiction
Alcohol related
Anthropometrics
Behavior
Birth defects
Cancer
Cancer treatment
Cardiovascular
Demographics
Diabetes type 1 or 2
Diabetic complications
Environmental exposures
Family history
Gastrointestinal
Gynecological/obstetric
Imaging
Infectious diseases and immunity
Laboratory
Mental health and cognition
Neurological
Nutrition, diet and dietary supplements
Occuplation/ocupational history
Ocular (eye)
Oral/dental
Pediatric
Physical activity and exercise
Preterm birth
Psychiatric
Psychosocial
Reproductive
Respiratory/pulmonary function
Skin, bone, muscle and joint
Speech and hearing
Substance use and dependence
Tobacco
Other
Phenotype data

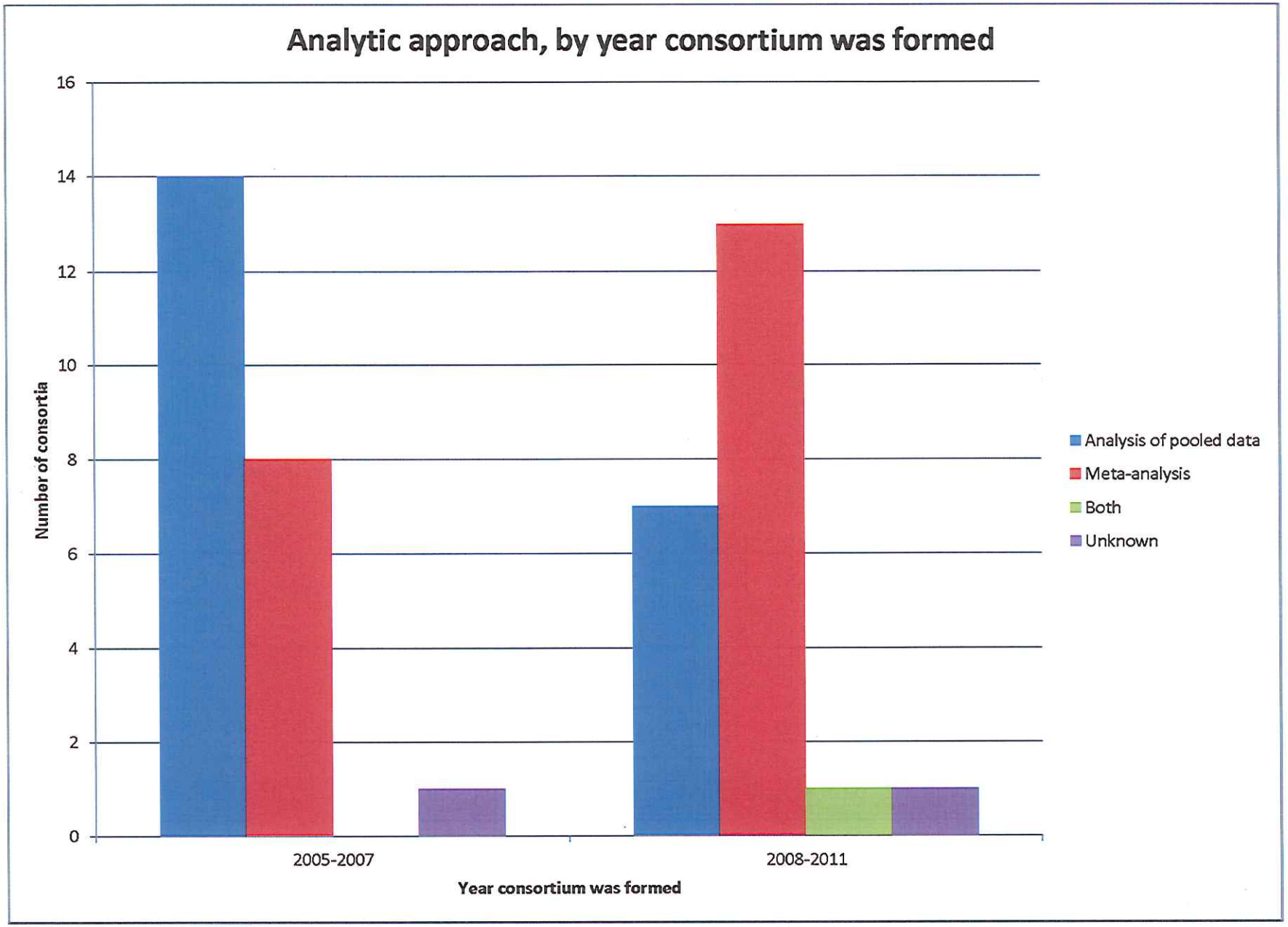

Consortium funding by number of contributing studies
Number of consortia
0
1
2
3
4
5
6
7
8
With funding (any type)
Without funding
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
Number of individual studies contributing to consortium


Race distribution by year consortium was formed
25
20
15
10
5
0
Number of consortia
1996-2004
2005-2007
2008-2011
Year consortium was formed
White
Asian
Black
Pacific Islander or Native Hawaiian
American Indian or Alaskan Native


Consortia subjects by age group
50
40
30
20
10
0
Number of consortia
Infants
Children
Teenagers
Young Adults
Middle Aged
Elderly
Age group


Consortia subjects by gender
50
40
30
20
10
0
Number of consortia
Women only
Both men and women
Gender


Genotyping platform by year consortium was formed
Number of consortia
0
2
4
6
8
10
12
14
16
18
1996-2004
2005-2007
2008-2011
Year consortium was formed
Affymetrix only
Illumina only
Both


Analytic approach, by year consortium was formed
16
14
12
10
8
6
4
2
0
Number of consortia
2005-2007
2008-2011
Year consortium was formed
Analysis of pooled data
Meta-analysis
Both
Unknown

Legends:

Figure 1. Size of consortia by number of subjects

This box and whisker plot shows size of consortia as defined by total number of subjects (y-axis) over 3 time periods based on the year of formation of the consortia (x-axis). The 2 different blocks in each of the 3 time periods represent the 2nd and 3rd quartiles; the horizontal line where the inner quartiles meet represents the median. The whiskers show the minimum and maximum number of subjects for each time period. The data are derived from the Consortia Survey.

Figure 2. Size of consortia by the number of participating GWAS

This box and whisker plot shows size of consortia as defined by number of contributing GWAS (y-axis) over 3 time periods based on the year of formation of the consortia (x-axis). The 2 different blocks represent the 2nd and 3rd quartiles; the horizontal line where the inner quartiles meet represents the median. The whiskers show the minimum and maximum number of contributing GWAS for each time period. The data are derived from the Consortia Survey.

Figure 3. Primary phenotype

Consortia were asked to indicate their primary phenotype or phenotypes of interest. These phenotypes are shown on the y-axis, the number of consortia are shown on the x-axis. The 'cancer' phenotype is the most frequently studied phenotype. The horizontal bar for the 'cancer' phenotype has varying colors that correspond to different types of primary cancer. The data are derived from the Consortia Survey.

Figure 4. GWAS with additional phenotype data, by category

GWAS investigators were asked to indicate what phenotype data were collected as part of their GWAS, beyond that of their primary phenotype or phenotypes. The categories are shown on the x-axis. Each GWAS investigator could report as many as 30 different categories of phenotype data (range 0-30). The data are derived from the GWAS Investigator Survey.

Supplemental Figure 1. Consortium funding by number of contributing studies

The number of consortia (y-axis) is shown in relation to the number of individual studies contributing to the consortium (x-axis). The different shades depict whether a consortium has funding or not based on the response to the Consortia Survey.

Supplemental Figure 2. Race distribution by year consortium was formed

The number of consortia (y-axis) that include subjects from various ancestry groups (x-axis) according to the year the consortium was formed. The data are derived from the Consortia Survey; respondents could indicate more than one group.

Supplemental Figure 3. Consortia subjects by age group

Age group is shown on the x-axis. The number of consortia addressing the age group indicated is depicted on the y-axis. The data is derived from the Consortia Survey; respondents could indicate more than one age group.

Supplemental Figure 4. Consortia subjects by gender

The number of consortia (y-axis) that include only women and those that include both women and men (x-axis). The data are from the Consortia Survey.

Supplemental Figure 5 Genotyping platform by year consortium was formed

The number of consortia (y-axis) according to the genotyping platform used by the consortium's participating GWAS, in t time periods (x-axis) representing the year the consortium was formed. As reported in the responses to the Consortia Survey.

Supplemental Figure 6. Analytic approach, by year consortium was formed

The number of consortia (y-axis) according to their analytic approach—analysis of pooled data, meta-analysis of summary data, or both—in two time periods (x-axis) based on the year the consortium was formed. As reported in the responses to the Consortia Survey.

Supplemental Table 1. Summary list of consortia from all sources

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
1	ABC (African-American Breast Cancer Consortium)	http://epi.grants.cancer.gov/Consortia/single/abc.html	Eight studies of women of African descent with and without breast cancer, supported by the National Cancer Institute and diverse other grants	Yes
2	ADGC (Alzheimer's Disease Genetic Consortium)	http://alois.med.upenn.edu/adgc/	An $18.3 million five-year research grant to a GWAS for late-onset Alzheimer's disease funded the National Institute on Aging	Yes
3	ANZgene (Australia and New Zealand Multiple Sclerosis Genetics Consortium)	http://www.msra.org.au/anzgene	Australian national collaborative effort to identify genes associated with multiple sclerosis	Yes
4	arcOGEN	www.arcogen.org.uk	UK-wide collaborative effort aiming to scan genome-wide over 7500 osteoarthritis cases in a two-stage genome-wide association scan	Yes
5	Asian Barrett's Consortium	http://epi.grants.cancer.gov/Consortia/single/asianbc.html	Focus on Barrett's and esophageal carcinoma in Asia	Not at this time
6	Asian Cohort Consortium	http://www.fhcrc.org/science/acc/; http://www.asiacohort.org/Pages/Default.aspx	Cohorts involving Asian studies and focusing on environment, genetics and populations	Not at this time
7	Attention Deficit Hyperactivity Disorder GWAS Consortium (Psychiatric GWAS Consortium)	http://pgc.unc.edu/	Goal is to conduct analyses of GWAS data for Attention Deficit Hyperactivity Disorder, autism, bipolar disorder, major depression disorder, and schizophrenia, with aims including: (1) Disorder-specific meta-analyses, (2) Cross-disorder analyses, and (3) Comorbidity meta-analyses.	Yes
8	BC2OS (Breast Cancer Consortium for Outcomes And Survival)	http://epi.grants.cancer.gov/Consortia/tables/breast.html#bc2os	Collaborative investigation led by investigators Betty Caan (Kaiser Permanente) and Xiao Ou Shu, M.D., Ph.D. Vanderbilt University Medical Center	Yes
9	BCAC (Breast Cancer Association Consortium)	http://www.srl.cam.ac.uk/consortia/bcac/index.html	Breast cancer investigators supported by Cancer Research UK (CRUK), the European Cooperation in Science and Technology (COST) and an EU FP7 Grant	Yes
10	B-CFR (Breast Cancer Family Registry)	http://epi.grants.cancer.gov/CFR/; http://www.ncbi.nlm.nih.gov/pmc/articles/PMC468645/	A collaboration of six academic and research institutions and their medical affiliates in the USA, Canada, and Australia	Yes
11	BEACON (International Barrett's and Esophageal Adenocarcinoma Consortium)	http://bea.tlvcloud.org/tag/beagess/	An international group of over three dozen investigators with completed or ongoing epidemiologic studies of esophageal adenocarcinoma and/or Barrett's esophagus	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
12	Body Mass Index (BMI) and All Cause Mortality Pooling Project	http://epi.grants.cancer.gov/bmi/participants.html	A collaborative effort among more than 20 cohorts from the National Cancer Institute's (NCI) Cohort Consortium to quantify the risk associated with being overweight or obese and the relationship between BMI and all-cause mortality varies by other factors. This Project is systematically analyzing pooled data from prospective studies encompassing 1.46 million adults.	Yes
13	BPC3 (Breast and Prostate Cancer Cohort Consortium)	http://epi.grants.cancer.gov/BPC3/	Investigators pool data and biospecimens from 10 large prospective cohorts to conduct research on gene-environment interactions in cancer etiology	Yes
14	BTEC (Brain Tumor Epidemiology Consortium)	http://epi.grants.cancer.gov/btec/	An open scientific forum organized to foster the development of multi-center, international and inter-disciplinary collaborations that will lead to a better understanding of the etiology, outcomes, and prevention of brain tumors	Yes
15	CADISP (Cervical Artery Dissections and Ischemic Stroke Patients)	http://www.chazard.org/cadisp	A European Consortium performing research on ischemic stroke in the young and in particular on cervical artery dissection (the most common cause of ischemic stroke in young people)	Yes
16	CALiCo Consortium (Genetic Epidemiology of Causal Variants Across the Life Course)	https://www.pagestudy.org/index.php/studies/58-calico	A consortium of well characterized population based studies and a central genotyping and resequencing core laboratory, to accelerate the understanding of the role and population impact of putative causal genetic variants related to complex diseases	Not at this time
17	CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis consortium)	http://www.ncbi.nlm.nih.gov/pubmed/20923989	CARDIoGRAM combines data from European GWAS including samples from: Atherosclerotic Disease VAscular functioN and genetiC Epidemiology study, CADomics, Cohorts for Heart and Aging Research in Genomic Epidemiology, deCODE, the German Myocardial Infarction Family Studies I, II, and III, Ludwigshafen Risk and Cardiovascular Heath Study/AtheroRemo, MedStar, Myocardial Infarction Genetics Consortium, Ottawa Heart Genomics Study, PennCath, and the Wellcome Trust Case Control Consortium	Yes
18	CARe (Candidate-gene Association REsource)	http://www.broadinstitute.org/gen_analysis/care/index.php/Main_Page	An NHLBI collaboration of up to 50,000 participants from nine cohort studies whose outputs will be analytic results from statistical and computational methods used to perform large-scale candidate gene association studies of phenotypes across multiple cohorts, whole-genome association studies, and tests for gene-gene and gene-environment interactions	Yes
19	C-CFR (Colon Cancer Family Registry)	http:// epi.grants.cancer.gov/CFR/about_colon.html	C-CFR is an international consortium of six research institutions and an Informatics Center. The participating centers are at the Fred Hutchinson Cancer Research Center; Mayo Clinic; University of Southern California; University of Queensland; Cancer Care Ontario,; and the University of Hawaii Cancer Research Center	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
20	C-CFR/GECCO (Colorectal Cancer GWAS Consortium)	http://epi.grants.cancer.gov/Consortia/tables/colorectal.html	An EGRP supported Epidemiology initiative	Yes
21	CGASP (Consortium of Genetic Association of Smoking Related Phenotypes)	http://www.nida.nih.gov/DirReports/DirRep510/DirectorReport16.html	A consortium sponsored by NIDA (National Institute on Drug Abuse) to conduct a meta-analysis as well as integrate data on the genetics of nicotine addiction, lung cancer, and COPD	Not at this time
22	CGN (Cancer Genetics Network)	http://www.cancergen.org/	A national network of centers specializing in the study of inherited predisposition to cancer. The CGN was formed in 1999, and receives funding from the National Cancer Institute. Includes a Data Coordinating Center, 14 Clinical Centers and an Informatics Group	Yes
23	CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology)	http://web.chargeconsortium.com	Formed to facilitate genome-wide association study meta-analyses and replication opportunities among multiple large and well-phenotyped longitudinal cohort studies	Yes
24	CIMBA (Consortium of Investigators of Modifiers of BRCA1/2)	http://www.srl.cam.ac.uk/consortia/cimba/index.html	A collaborative group of researchers working on genetic modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers. The aim is to provide sufficient sample sizes to allow large scale studies in order to evaluate reliably the effects of genetic modifiers	Yes
25	CKDGen Consortium	http://www.jhsph.edu/publichealthnews/press_releases/2010/kao_kidney_risk_genes.html	A team of researchers from the United States and Europe investigating the role of genes in the etiology of common forms of kidney disease	Yes
26	CLIC (Childhood Leukemia International Consortium)	https://ccls.berkeley.edu/clic/	An international group of researchers established in 2006 in collaboration with the University of California at Berkeley and the International Agency for Research on Cancer, with the aim of sharing comparable data from case-control and family-based trio studies on childhood leukemia. To coordinate scientific efforts and achieve greater power to elucidate the role of a variety of clinical, infectious, environmental, and genetic risk factors in the etiology of childhood leukemia, especially for rare subtypes	Yes
27	COGENT (COlorectal cancer GENeTics)	[no website link found]	An international consortium to study the role of polymorphic variation on the risk of colorectal cancer	Yes
28	Cohort Consortium	http://epi.grants.cancer.gov/Consortia/cohort.html	An extramural-intramural partnership formed by the National Cancer Institute (NCI) to address the need for large-scale collaborations to pool the large quantity of data and biospecimens necessary to conduct a wide range of cancer studies, involves over 40 studies representing > 4 million persons	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
29	CRC (Chronic Lymphocytic Leukemia Research Consortium)	http://mayoresearch.mayo.edu/slager_lab/gec.cfm	A family based, multi-center, multidisciplinary consortium funded by the National Cancer Institute (U01 CA 118444) and including investigators from Mayo Clinic, M.D. Anderson Cancer Center, National Cancer Institute, University of California in San Diego, University of Minnesota, Duke University, and the University of Utah	Yes
30	DentalSCORE (Dental Strategies Concentrating on Risk Evaluation)	[no website link found]	A collaborative study adding an oral and dental exam component to HeartSCORE, a community-based longitudinal study focusing on racial and socioeconomic disparities in cardiovascular risk	Not at this time
31	DGI (Diabetes Genetics Initiative)	http://www.broad.mit.edu/diabetes	A collaboration of the Broad Institute of MIT and Harvard, Lund University, and Novartis Institutes for BioMedical Research to identify the genetic determinants of type 2 diabetes. This collaboration aims to collect and analyze samples from type 2 diabetic patients from nations across the globe, performing whole genome scans to provide a comprehensive view of the DNA sequence variants associated with the disease	Yes
32	DIAGRAM (Diabetes Genetics Replication And Meta-analysis Consortium)	http://www.well.ox.ac.uk/DIAGRAM/index.html	A genome-wide association study of type 2 diabetes (T2D) from the FUSION, DGI, and WTCCC/UKT2D groups	Yes
33	DRIVE (Discovery, Biology, and Risk of inherited Variants in breast Cancer	http://cancercontrol.cancer.gov/grants/abstract.asp?ApplID=7861271	A post GWAS study to investigate genetic variants associated with breast cancer. PI is David Hunter at Harvard.	Yes
34	E2C2 (Epidemiology of Endometrial Cancer Consortium)	epi.grants.cancer.gov/e2c2; http://www.ncbi.nlm.nih.gov/pubmed/19132539	A consortia formed in 2006 to pool data from existing studies. This allows for investigations of uncommon risk factors, risk for rare histologic subtypes, and associations within strata that cannot be achieved in individual studies	Yes
35	EADI (European Alzheimer's Disease Initiative)	http://alz.org/documents_custom/FINAL_AlzAssoc_IGAP_news_release_1-31-11.pdf	Four groups are represented: 1) The European Alzheimer's Disease Initiative (EADI) in France led by Philippe Amouyel at the Institute Pasteur de Lille and Lille University, 2) The Alzheimer's Disease Genetics Consortium (ADGC) from the USA led by Gerard Schellenberg, Ph.D. at the University of Pennsylvania School of Medicine, 3) The Genetic and Environmental Risk in Alzheimer's Disease (GERAD) from the UK led by Julie Williams at Cardiff University, 4) the neurology subgroup of the Cohorts for Heart and Aging in Genomic Epidemiology (CHARGE) led by Sudha Seshadri at Boston University	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
36	EGG (Early Growth Genetics)	http://www.well.ox.ac.uk/mccarthy-2	One of a group of major international consortia including the Wellcome Trust Case Control Consortium (WTCCC, genome-wide association studies), DIAGRAM (T2D genetics), EU-funded ENGAGE consortium (www.euengage.org), GIANT (anthropometric traits), MAGIC (continuous glycaemic traits), EGG (early growth genetics), and International 1q Consortium.	Yes
37	eMERGE (Electronic Medical Records & Genomics)	https://www.mc.vanderbilt.edu/victr/dcc/projects/acc	A national consortium formed to combine DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research	Yes
38	ENGAGE (European Network of Genomic and Genetic Epidemiology)	http://www.euengage.org	A 5-year research project started in January 2008 and funded with 12 million Euros by the European Commission. The ENGAGE Consortium includes 23 research organizations and two biotechnology and pharmaceutical companies across Europe and in Canada and Australia. Its goal is to identify large numbers of novel susceptibility genes that influence metabolic, behavioral and cardiovascular traits, and to study the interactions between genes and life style factors.	Yes
39	ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles Consortium)	http://enigmaconsortium.org/	A collaboration focused on study of unclassified variants in BRCA1 and BRCA2 tumor suppressor genes, in predisposition to breast and ovarian cancer, through collaborative large-scale projects by sharing data and improving classification methods. Co-ordination of ENIGMA is funded by The National Institutes of Health Recovery Act supplement award (CA116167Z)	Not at this time
40	EUROCRAN (European Collaboration on Craniofacial Anomalies)	http://www.eurocran.org/default.asp	Funded by the European Commission (EC) under the Quality of Life theme of Framework Programme V, EUROCRAN involves investigators from 19 European centers with the aim of study of craniofacial anomalies (CFA)	Yes
41	EUROSPAN (European Special Population Network)	http://homepages.ed.ac.uk/s0565445/	Project to investigate 5 genetically isolated European populations	Yes
42	GAPPS (Global Alliance to Prevent Prematurity and Stillbirth)	http://gappsseattle.org	GAPPS collaborates to help catalyze research to understand causes and find solutions to preterm delivery. GAPPS is developing a repository of data and specimens from diverse women to provide a resource for researchers around the world	Yes
43	GARNET (Genomics and Randomized Trials Network, or Genome-wide Association Research Network into Effects of Treatment)	www.garnetstudy.org	A series of genome-wide association studies of treatment response in randomized clinical trials that looks to identify genetic variants associated with response to treatments for conditions of clinical or public health significance	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
44	GEC (Genetic Epidemiology of Chronic Lymphocytic Leukemia Consortium)	http://mayoresearch.mayo.edu/slager_lab/gec.cfm	A multi-center, multidisciplinary consortium to specifically investigate the genetic basis of CLL through the use of high-risk CLL families. Funded by the National Cancer Institute (U01 CA 118444) and including investigators from Mayo Clinic, M.D. Anderson Cancer Center, National Cancer Institute, University of California in San Diego, University of Minnesota, Duke University, and the University of Utah	Yes
45	GECCO (Colorectal Cancer GWAS Consortium)	http://epi.grants.cancer.gov/Consortia/tables/colorectal.html#gecco	To accelerate the discovery of colorectal cancer-related variants by (1) conducting a pooled analysis of all five existing colorectal cancer GWAS and validating findings from this pooled analysis in a large-scale replication study; (2) establishing the identity of the underlying causal variants in these true genetic regions; and (3) investigating potentially important interactions between genetic variants and environmental factors	Yes
46	GEFOS (Genetic Factors of Osteoporosis Consortium)	http://www.gefos.org	A large international collaboration that proposes to capitalize on the success of GENOMOS by using Genome Wide Association (GWA) analysis with high density SNP arrays to identify common risk gene variants for osteoporosis	Yes
47	GELCC (Genetic Epidemiology of Lung Cancer)	http://www.eh.uc.edu/gelcc/; http://epi.grants.cancer.gov/Consortia/single/gelcc.html	A consortium of investigators investigating familial lung cancer	Yes
48	GEM (Genes, Environment and Melanoma)	http://gemstudy.org/main/index.html	A population-based international consortium studying risk for melanoma development and progression. The representative GEM populations currently consist of approximately 3700 individuals from the United States (California, New Jersey, North Carolina), Canada (British Columbia and Ontario), Italy (Turin), and Australia (New South Wales and Tasmania).	Not at this time
49	GEMINI (Gene-Environment Meta-analysis of Nicotine dependence) Consortium	http://zork.wustl.edu/cogend/	A consortium of adolescent and young adult studies (up to age 24) with the aim of carrying out a meta-analysis of genetic factors for nicotine dependence and extensive smoking in adolescence	Yes
50	GENEVA (GENe EnVironment Association studies)	http://www.genevastudy.org	The genetics component of an NIH-wide initiative that aims to accelerate understanding of genetic and environmental contributions to health and disease, its aims are to identify genetic variants related to common, complex diseases, identify variations in gene-trait associations related to environmental exposures, and address potential pathways to disparities in health outcome	Yes
51	GenoMEL (Melanoma Genetics Consortium)	[no website link found]		Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
52	GIANT (Genome-wide Investigation of ANThropometric measures)	http://www.helmholtz-muenchen.de/epi/beitraege-zu-netzwerken/giant-genomewide-investigation-of-anthropometric-measures/index.html	An international assembly of study investigators and genetic epidemiologists which was established to pool genome-wide association results on anthropometric parameters such as body-mass-index and height. The GIANT consortium has the aim to determine all relevant genes that are involved in modulating weight, height and more complex measures of obesity	Yes
53	GLIOGENE	www.gliogene.org	An international consortia to understand familial glioma	Yes
54	Global BPGen Consortium	[no website link found]	A multinational collaboration with investigators from across the US and Europe that seeks to identify genetic variants in known and novel genes that influence blood pressure in the general population	Yes
55	Global Lipid Genetics Consortium	http://communications.uwo.ca/com/western_news/stories/hegele_part_of_worldwide_gene,_cholesterol_study_20100804446648/ (LINK to article referencing study by the global lipid genetics consortium); http://docs.google.com/viewer?a=v&q=cache:YUVd7RBM8OEJ:www.ingelsson.org/Global%2520Lipids%2520Genetics%2520Consortium.pps+global+lipid+genetics+consortium&hl=en&gl=us&pid=bl&srcid=ADGEESgy1Bk5NJA8lkKQzkuC4jobX2S5F6MXKEQD-81qnnvuZg8IgbtRwRjva_tnRTeBaam43ksfapAQZ7lioE_YLHGxcwM6OiV-vi3EhxNNqki57jpuKfilqZaAyBDcmmRg3REuMgcX&sig=AHIEtbT0LGo8cBKByGixQF2Br_fD6z79qw	The aim of this consortium is to study the genetic determinants of blood low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol and triglycerides	Yes
56	I4C (International Childhood Cancer Cohort Consortium)	http://www.cls.ioe.ac.uk/text.asp?section=00010001000500090018	A consortium of studies of children/adolescents referred to as the *International Childhood Cancer Cohort Consortium* (I4C) involving primary investigators of diverse cohorts including (NCS, TIHS, DNBC, and the China Family and Children Cohort Study	Yes
57	ICBC (International Consortium of Bladder Cancer)	http://icbc.cancer.gov	Formed in 2005, the International Consortium of Bladder Cancer (ICBC) was designed as an open scientific forum for epidemiologic research in bladder cancer.	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
58	ICPCG (International Consortium on Prostate Cancer Genetics)	https://www.icpcg.org/	Established in 1996, a consortium involving ~2000 prostate cancer families (including over 110 African American PCa families), with special emphasis on subsets of families with early age at diagnosis, large numbers of affected individuals, and multiple cases with features of clinically aggressive disease	Yes
59	IGAP (International Genomics of Alzheimer's Project)	http://www.alzheimersreadingroom.com/2011/02/international-genomics-of-alzheimers.html	A collaboration involving 4 groups in the USA and Europe: - The European Alzheimer's Disease Initiative (EADI) in France led by Philippe Amouyel, at the Institute Pasteur de Lille and Lille University; - The Alzheimer's Disease Genetics Consortium (ADGC) from the USA led by Gerard Schellenberg, at the University of Pennsylvania School of Medicine; - The Genetic and Environmental Risk in Alzheimer's Disease (GERAD) from the UK led by Julie Williams, Ph.D., at Cardiff University; - The neurology subgroup of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) led by Sudha Seshadri, at Boston University.	Yes
60	IIBDGC (International Inflammatory Bowel Disease Genetics Consortium)	http://medicine.yale.edu/intmed/ibdgc/resources/index.aspx	A collaborative effort of NIDDK IBDGC including six Genetic Research Centers as well as with members of the IBD and genetics communities. Contact is Yashoda Sharma, PhD, Project Administrator, IBDGC	Yes
61	ILCCO (International Lung Cancer Consortium)	http://ilcco.iarc.fr	An international group of lung cancer researchers established in 2004 with the aim of sharing comparable data from ongoing lung cancer case-control and cohort studies from different geographical areas and ethnicities	Yes
62	IMMC (International Multiple Myeloma Consortium)	http://epi.grants.cancer.gov/Consortia/single/immc.html	A consortium of North American investigators interested in multiple myeloma	Not at this time
63	INHANCE (International Head and Neck Cancer Consortium)	http://inhance.iarc.fr	Established in 2004, a collaboration of research groups leading large molecular epidemiology studies of head & neck cancer that are on-going or have been recently completed	Yes
64	InterLymph (International Consortium on Lymphoma Epidemiologic Studies)	http://epi.grants.cancer.gov/InterLymph/	Formed in 2001, the Consortium is a group of international investigators who have completed or have ongoing case-control studies and who discuss and undertake research projects that pool data across studies or otherwise undertake collaborative research. The InterLymph Consortium with its five working groups conducts annual meetings to discuss, coordinate and develop ongoing and future research projects	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
65	International Cancer Genome Consortium	http://www.icgc.org/	Goal is to obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe	Not at this time
66	International Endogene Consortium	http://endometriosis.org/news/research/genome-wide-association-study-identifies-dna-predisposing-women-to-endometriosis/ (consortium website)	Established in 2008, the International Endogene Consortium is a collaborative effort between investigators in the USA, the UK and Australia. The goal is to identify genetic determinants of various stages of endometriosis.	Yes
67	International Headache Genetics Consortium	http://www.headachegenetics.org/	An international collaboration between research teams interested in mechanisms and pathophysiology of headache and related disorders consisting of groups from Finland, Germany, the Netherlands, Spain and the UK. Chaired by prof. Aarno Palotie at the Sanger Institute members@headachegenetics.org	Yes
68	International Kawasaki Disease Genetics Consortium	http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1000319	Kawasaki Disease cases and families identified by Dutch, UK and US and Australian investigators	Yes
69	International Type 2 Diabetes Consortium	http://csg.sph.umich.edu/consortium	A consortium of groups mapping genes for NIDDM in diverse populations that has come together to carry out a joint analysis of their linkage data	Yes
70	ISGC (International Stroke Genetics Consortium)	http://www.strokegenetics.org	A consortium formed in early 2007 for the purpose of facilitating collaborative efforts to perform large scale, multi-center genetic studies in stroke. The aims include: 1) assembling a large well-characterized sample of stroke subjects with DNA, all of whom have been carefully phenotyped; and 2) harmonizing phenotype information to allow easy and reliable primary and secondary analyses of the combined cohort	Yes
71	LACE (Latin American Cancer Epidemiology Consortium)	http://epi.grants.cancer.gov/Consortia/single/lace.html	No information available	Yes
72	LIPID MAPS (Lipid Metabolites And Pathways Strategy)	www.lipidmaps.org	A multi-institutional effort created in 2003 to identify and quantitate, using a systems biology approach and sophisticated mass spectrometers, all of the major and many minor lipid species in mammalian cells, as well as to quantitate the changes in these species in response to perturbation. The ultimate goal is to better understand lipid metabolism and the active role lipids play in diabetes, stroke, cancer, arthritis, Alzheimer's and other lipid-based diseases in order to facilitate development of more effective treatments.	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
73	MADCaP (Men of African Descent and Prostate Cancer)	http://epi.grants.cancer.gov/madcap/index.html	A consortium formed in 2007 including investigators from the US, UK, Africa and the Caribbean, who have cohort or case-control studies on prostate cancer among men of African ancestry. The investigators share common interests in studying health disparities, underserved populations, and the role that genetic and non-genetic factors play in the disparities. An aim is to validate and extend findings on genomic regions that have been associated with prostate cancer risk in genome-wide association studies (GWAS) using a large sample of participants from consortium members	Yes
74	MAGIC (The Meta-Analyses of Glucose and Insulin-related traits Consortium)	http://www.magicinvestigators.org/	A collaborative effort to combine data from multiple GWAS to identify loci that affect glycemic and metabolic traits. The genetic studies of fasting glucose levels originated as four distinct consortia: (i) European Network for Genetic and Genomic Epidemiology (ENGAGE), combining data from deCODE, Northern Finland Birth Cohort 1966 (NFBC1966), Netherlands Twins Register/Netherlands Study of Depression and Anxiety (NTR/NESDA) and the Rotterdam Study; (ii) Genetics of Energy Metabolism (GEM), a meta-analysis of the Lausanne (CoLaus) and TwinsUK scans; (iii) DFS, involving the Diabetes Genetics Initiative (DGI), Finland-United States Investigation of NIDDM Genetics (FUSION) and SardiNIA scans; and (iv) the Framingham Heart Study (FHS)	Yes
75	MECC (Molecular Epidemiology of Colorectal Cancer)	http://sitemaker.umich.edu/gruber.lab/gruber.research/mecc	Established in 1998, The University of Michigan and The National Israeli Cancer Control Center collaborated to produce this population based case-control study of colorectal cancer	Yes
76	NEIGHBOR (National Eye Institute Glaucoma Human Genetics CollaBORation)	http://www.nei.nih.gov/news/statements/glaucoma_initiatives.asp	A consortium aiming to identify genetic determinants for primary open angle glaucoma (POAG)	Yes
77	neuroCHARGE: CHARGE neurology working group	http://depts.washington.edu/chargeco/wiki/Main_Page	This is a subset of the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) consortium with a focus on neurologic phenotypes	Yes
78	NGFN (German National Genome Research Network)	http://www.ngfn.de/en/	An endeavor to research diseases with high incidence in Germany including cancer, cardiovascular diseases, diseases of the nervous system, infections and inflammation as well as diseases linked to environmental factors	Yes
79	Nutritional Genomics Consortium	http://sites.channing.harvard.edu/nutritionalgenomics	[no information found]	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
80	OCAC (Ovarian Cancer Association Consortium)	http://www.srl.cam.ac.uk/consortia/ocac/index.html	An international multidisciplinary consortium Formed in April 2005 that meets biannually, it is forum of investigators of case-control studies of ovarian cancer and currently is the only functional venue for large-scale validation of ovarian cancer susceptibility polymorphisms, Its aim is to combine data from many studies, to provide a reliable assessment of the risks associated with these genes.	Yes
81	OCCC (Ovarian Cancer Cohort Consortium)	[no website link found]	[no information found]	Yes
82	Ox-GSK (Oxford-GlaxoSmithKline consortium)	[no website link found]	The Ox-GSK consortia has been referenced by several works, often in conjunction w/ various other consortia in tobacco studies (see, TAG, Tobacco and genetics Consortium)	Yes
83	P3G Consortium (Public Population Project in Genomics)	http://www.p3g.org	A non-profit international consortium founded in 2003 to respond to the growing needs and demands of the population genomics community, its goal is to foster harmonization of research tools developed by its members	Yes
84	PACGENE (Pancreatic Cancer Genetic Epidemiology Study)	http://clinicaltrials.gov/ct2/show/NCT00526578	A multicenter, multidisciplinary Pancreatic Cancer Genetic Epidemiology (PACGENE) consortium to identify susceptibility genes in high risk FPC pedigrees using cutting-edge genetic analysis methods	Yes
85	PAGE (Population Architecture using Genomics and Epidemiology)	https://www.pagestudy.org	The purpose of this program is to provide support for the investigation, in well-characterized population studies, of genetic variants identified as potentially causally associated with complex diseases in genome-wide association (GWA) and other genetic studies, with the aim of widespread sharing of the resulting population-based descriptive and association data to accelerate the understanding of genes related to complex diseases	Not at this time
86	PANC4 (Pancreatic Cancer Case Control Consortium)	http://panc4.org/	Established in 2006, the PANC4 consortium is a collaborative, interdisciplinary, international effort to further scientific understanding of the cause of pancreatic cancer through analyses of pooled data. The PANC4 consortia represents an expansion of the PanScan study to include eight hospital-based case-control studies.	Yes
87	Pancreatic Cancer Cohort Consortium	http://epi.grants.cancer.gov/PanScan/	Consisting of more than a dozen prospective epidemiologic cohort studies within the larger NCI Cohort consortium, the Pancreatic Cancer Cohort Consortium was formed in 2006 to identify genetic factors, environmental exposures, and gene-environment interactions that contribute to the development of pancreatic cancer, with an emphasis on facilitating early detection of the disease.	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
88	PanScan	http://epi.grants.cancer.gov/PanScan/	The PanScan consortia was formed in 2006 to facilitate two genome-wide association studies, PanScan I and PanScan II, which have led to the discovery of four novel regions in the genome associated with risk for pancreatic cancer.	Yes
89	PBTC (Pediatric Brain Tumor Consortium)	http://www.pbtc.org/	Formed by the National Cancer Institute in 1999, the aim of this consortium is to study correlative tumor biology and investigate new therapies for primary childhood CNS tumors.	Yes
90	PGRN Pharmacogenomics Research Network (PGRN)	http://www.nigms.nih.gov/Research/FeaturedPrograms/PGRN/	Involves diverse and ongoing collaborations between the NIH Pharmacogenomics Research Network and PharmGKB, as well as other researchers working in pharmacogenomics	Yes
91	POCRC (Pacific Ovarian Cancer Research Consortium)	http://www.pocrc.org/newsletters/POCRCNews-2010.pdf	SPORE-funded, clinically and biomarker based investigations in ovarian cancer	Not at this time
92	Post-Genome Wide Association Initiative	http://epi.grants.cancer.gov/pgwas/	The National Cancer Institute has awarded five cooperative agreements for transdisciplinary research projects to exploit findings from existing genome-wide association studies (GWAS) and accelerate new discoveries (see DRIVE, ELLIPSE, FOCI, TRICL, CORECT)	Yes
93	PREGENIA (Preterm Birth and Genetics International Alliances)	http://www.prebic.net/documents/MOA%20final%209-5-07.pdf	A consortium established to conduct research to understand genetic predisposition in preterm birth, a leading cause of neonatal morbidity and mortality. The aim is to better understand the genetic basis of preterm birth, and to develop genetic tools to predict women who are at risk for having preterm labor	Yes
94	PROSE (Prevention and Observation of Women of African Ancestry Breast Cancer Study Surgical Endpoints)	http://www.cceb.upenn.edu/pages/prose/index.html	Its goal is to estimate breast and ovarian cancer risk reduction after the use of risk-reduction surgery and to evaluate psychosocial endpoints in women who carry BRCA1 and BRCA2 mutations.	Yes
95	Radiogenomics Consortium	http://www.ncbi.nlm.nih.gov/pubmed/20338472	Established in Manchester, U K, 2009, in conjunction with the 15th L. H. Gray Workshop with investigators from throughout the world. The unifying interest of these researchers is the identification of genetic variants, primarily single nucleotide polymorphisms (SNPs), associated with the development of normal tissue toxicities from radiation therapy. Goals are to develop an assay capable of predicting which cancer patients are most likely to develop radiation injuries resulting from treatment and to elucidate the molecular pathways responsible for radiation-induced toxicities	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
96	ReproGen Consortium	http://hmg.oxfordjournals.org/content/suppl/2010/09/28/ddq417.DC1/ddq417supp.pdf [this is a link to a summary of the collaborators and research results]	A collaborative effort involving 175 researchers from 104 international institutions, including researchers at the Boston University School of Medicine and Boston University School of Public Health, the consortia aim is to identify genes controlling the age of sexual maturation in women.	Yes
97	SHARe (SNP Health Association Research)	http://public.nhlbi.nih.gov/GeneticsGenomics/home/share.aspx	Conducts NHLBI-funded genome wide association studies and analyses in several large NHLBI Cohort studies to identify genes underlying cardiovascular and lung disease and other disorders like osteoporosis and diabetes	Yes
98	SHARP (SHARe Asthma Resource Project)	http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000166.v1.p1	Conducts a genome-wide analysis in adults and children who have participated in NHLBI's clinical research trials on asthma	Yes
99	SpiroMeta Consortium	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3098839/	A consortium to facilitate large-scale meta-analysis of GWAS of lung function.	Yes
100	STAMPEED (SNP Typing for Association with Multiple Phenotypes from Existing Epidemiologic Data)	http://public.nhlbi.nih.gov/GeneticsGenomics/home/stampeed.aspx	Established in 2006, this consortia aims to identify genetic variants related to heart, lung, and blood disorders and their risk factors.	Yes
101	STOMP (Study of Tobacco in Minority Populations) Genetics Consortium	http://www.nida.nih.gov/about/organization/genetics/geneticsconsortia/index.php	STOMP was formed in 2010. It consists of multiple investigators from various studies who are interested in conducted a meta-analysis of GWAS data for smoking behavior among African-Americans.	Yes
102	SUNLIGHT Consortium (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits)	http://journals.lww.com/obgynsurvey/Abstract/2011/02000/Common_Genetic_Determinants_of_Vitamin_D.11.aspx (article)	A consortium of 15 epidemiologic studies from the U.S., U.K., Canada, Netherlands, Sweden and Finland investigating the genetic contribution to vitamin D status in almost 32,000 white individuals of European descent	Yes
103	TAG (The Tobacco, Alcohol and Genetics Consortium)	[no website link found]	Its goal is to conduct a genome-wide association study (GWAS) meta-analysis for smoking-related phenotypes using genotype and smoking phenotype data from existing GWAS of other traits e.g., heart disease, diabetes, etc. The sample size of the TAG Consortium may approach 100,000 subjects and is a multi-national collaboration which has the potential to identify novel genetic loci for smoking	Yes

	Consortium Name	Website URL, Identifier or Key article	Brief description	GWAS data available
104	TREAT-OA (Translational Research in Europe Applied Technologies for OsteoArthritis)	www.treatoa.eu	A large-scale collaborative, integrated, trans-disciplinary project to address the need for better treatment and diagnostics for osteoarthritis, the most common cause of disability in Europe, its goal is to identify diagnostic and prognostic genetic markers for disease risk and progression and potential therapeutic targets	Yes
105	TRICL (Transdisciplinary Research in Cancer of the Lung)	http://u19tricl.org/	The aim of this consortium is to identify sources of interindividual variability in lung cancer susceptibility involving large scale populations studies of lung cancer with GWAS data	Yes
106	Twins Eye Study in Tasmania and Brisbane Adolescent Twin Study	[no consortia website found]	[the only information found was particular to specific studies]	Yes
107	Type 1 Diabetes Genetics Consortium	https://www.t1dgc.org/home.cfm	The Type 1 Diabetes Genetics Consortium (T1DGC) is an international, multicenter program organized to promote research to identify genes and their alleles that determine an individual's risk for type 1 diabetes (T1D). The primary goal of the T1DGC is to establish resources and data that can be used by, and that is fully accessible to, the research community in the study of T1D	Yes
108	VDPP (Vitamin D Pooling Project)	http://epi.grants.cancer.gov/Vitamin D/	The VDPP has brought together researchers from 10 cohorts to conduct a large prospective epidemiologic study of the association between vitamin D status and 7 rarer cancers: endometrial, esophageal, stomach, ovarian, pancreatic, and kidney cancers, and non-Hodgkin lymphoma.	Yes
109	WECARE (Women, Cancer and Radiation Exposure)	http://skiweb.mskcc.org/WECARE/front.html	To investigate gene-environment interactions influencing susceptibility in the etiology of breast cancer by establishing a repository of epidemiologic risk factor information and biologic specimens from 700 women with asynchronous bilateral breast cancer and 1400 women with unilateral breast cancer who will be ascertained through 5 population-based tumor registries in the US and Denmark. Its plan is to examine the interaction of radiation exposure, the ATM gene, and breast cancer	Yes
110	WTCCC (Wellcome Trust Case-Control Consortium)	http://www.wtccc.org.uk	A group of 50 research groups across the UK established in 2005, the WTCCC aims are to exploit progress in understanding of patterns of human genome sequence variation along with advances in high-throughput genotyping technologies, and to explore the utility, design and analyses of genome-wide association (GWA) studies	Yes