

Editors' note: This series addresses topics that affect epidemiologists across a range of specialties. Commentaries are first invited as talks at symposia organized by the Editors. This paper was originally presented at the 2006 Congress of Epidemiology in Seattle.

The Emergence of Networks in Human Genome Epidemiology

Challenges and Opportunities

Daniela Seminara,* Muin J. Khoury,† Thomas R. O'Brien,‡ Teri Manolio,§ Marta L. Gwinn,† Julian Little,|| Julian P. T. Higgins,#¶ Jonine L. Bernstein,** Paolo Boffetta,†† Melissa Bondy,‡‡ Molly S. Bray,§§ Paul E. Brenchley,|||| Patricia A. Buffler,¶¶ Juan Pablo Casas,## Anand P. Chokalingam,*** John Danesh,††† George Davey Smith,‡‡‡ Siobhan Dolan,§§§ Ross Duncan,||||| Nelleke A. Gruis,¶¶¶ Mia Hashibe,†† David Hunter,¶¶¶ Marjo-Riitta Jarvelin,####* Beatrice Malmer,†††† Demetrius M. Maraganore,‡‡‡‡ Julia A. Newton-Bishop,§§§§ Elio Riboli,†† Georgia Salanti,¶ Emanuela Taioli,||||| Nic Timpson,††† André G. Uitterlinden,¶¶¶¶ Paolo Vineis,##### Nick Wareham,***** Deborah M. Winn,* Ron Zimmern,# and John P. A. Ioannidis††††‡‡‡‡ for the Human Genome Epidemiology Network and the Network of Investigator Networks

Large-scale “big science” is advocated as an approach to complex research problems in many scientific areas.¹ Epidemiologists have long recognized the value of large collaborative studies to address important questions that are beyond the scope of a study conducted at a single institution.² We define networks (or, interchangeably, consortia) as groups of scientists from multiple institutions who cooperate in research efforts involving, but not limited to, the conduct, analysis, and synthesis of information from multiple population studies. Networks, by virtue of their greater scope, resources, population size, and opportunities for interdisciplinary collaboration, can address complex scientific questions that a single team alone cannot.³

There is a strong rationale for using networks in human genome epidemiology particularly. Genetic epidemiology benefits from a large-scale population-based approach to identify genes underlying complex common diseases, to assess associations between genetic variants and disease susceptibility, and to examine potential gene–environment interactions.^{4–6} Because the epidemiologic risk for an individual genetic variant is likely to be small, a large sample size is needed for adequate statistical power.⁷ Power issues are even more pressing for less common disease outcomes. Replication in different populations and exposure settings is also required to confirm and validate results. The adoption of common guidelines for the conduct, analysis, reporting, and integration of studies across different teams is essential for credible replication. Transparency in acknowledging and incorporating both “positive” and “negative” results is necessary to direct subsequent research. Furthermore, newer and more efficient genotyping technologies must be integrated rapidly into current and planned population studies.^{8,9} Networks can support studies with sample sizes large enough to achieve “definitive” results, promote spinoff

Supported by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics and the National Cancer Institute Extramural Program, Division of Cancer Control and Population Sciences, Epidemiology and Genetics Research Program.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Editors' note: Related papers appear on pages 9, 13, and 18.

Correspondence: Daniela Seminara, Epidemiology and Genetics Research Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, NIH, EPN Bldg., Rm. 5142, MSC 7393, 6130 Executive Blvd., Bethesda, MD 20892. E-mail: seminard@mail.nih.gov.

Copyright © 2006 by Lippincott Williams & Wilkins

ISSN: 1044-3983/07/1801-0001

DOI: 10.1097/01.ede.0000249540.17855.b7

research projects, and yield faster “translation” of results into clinical and public health applications. Networks can also foster interdisciplinary and international collaboration.¹⁰ Lastly, networks can assemble databases that are useful for developing and applying new statistical methods for large data sets.¹¹

The experience of established networks provides an important knowledge base on which to develop recommendations for improving future efforts.¹² The Human Genome Epidemiology Network (HuGENet) recently launched a global network of consortia working on human genome epidemiology.¹³ This Network of Investigator Networks aims to create a resource to share information, to offer methodologic support, to generate inclusive overviews of studies conducted in specific fields, and to facilitate rapid confirmation of findings. In October 2005, HuGENet brought together representatives from established and emerging networks to share their experiences at a workshop in Cambridge, U.K.¹⁴ In advance of the meeting, a qualitative questionnaire was distributed to workshop participants. The questionnaire elicited information on experiences and practices in building and maintaining consortia. This article reports on the numerous challenges and their possible solutions as identified by the workshop participants (summarized in Table 1) as well as new opportunities offered by the network approach to genetic and genomic epidemiology.

SCIENTIFIC APPROACH

Selection of Scientific Questions

To date, most networks have targeted projects originating from preliminary evidence of specific associations or for the purpose of genetic linkage. In most consortia, projects are selected through group discussion and informal or semiformal (eg, voting) prioritization of candidate gene targets. Most networks try to focus on the best possible candidates to generate definitive evidence, but, given the large proportion of false-positives in genetic epidemiology,¹⁵ there is considerable uncertainty about the criteria for selecting such targets. Possible criteria include the number and consistency of published reports for a specific gene, the presence of a high-profile controversy in the literature, strong a priori biologic plausibility, potentially high population-attributable risk (eg, a common polymorphism) supporting linkage evidence from genomewide data, and candidates derived from genomewide association screens.^{16,17}

Networks are often focused on candidate genes involved in pathogenesis of the disease outcome or in biologic pathways involving environmental exposures such as metabolism of carcinogens.¹⁸ For example, the WECARE consortium on genetics of cancer and radiation exposure¹⁹ has addressed individual genes that lie within pathways related to double-strand breaks caused by radiation damage. Consortia are increasingly used to replicate findings from hypothesis-free genomewide approaches. For example, consortia are attempting to replicate findings from 2-stage genomewide association studies of Parkinson disease²⁰ and breast cancer.²¹ With decreasing genotyping cost and the expressed interest of funding agencies in genomewide associa-

tion studies,²² some consortia are coordinating large-scale genotyping and replication of whole genome association designs.²³

Prospective and Retrospective Components

Networks use information and biologic specimens from ongoing or established cohort and case-control studies with data on phenotypes. Phenotype information may have been accumulated either retrospectively or prospectively depending on the study design. Participating teams with prospective designs usually continue collecting phenotype information.

Regarding genotyping, several consortia perform meta-analyses of individual-level data using studies in which all genotyping has already been done and data have been published. Some consortia include additional genotyping from teams that have not yet done or published such genotyping; for other consortia, prospective genotyping represents the majority of the data. Increasingly, prospective genotyping is coordinated to test novel candidate gene variants or variants identified by genomewide approaches.

Handling of Information From Nonparticipating Teams

Many networks do not encompass all teams working on the disease or subject matter of interest. For some common diseases (eg, breast cancer), there are 2 or more organized multiteam consortia in addition to nonorganized teams.^{24–26} Some consortia attempt analyses that include outside data to examine the robustness of their findings. Integration of evidence across networks and across participating and nonparticipating teams remains a challenge in developing all-encompassing synopses of the evidence on specific gene-disease associations.²⁷

LAUNCHING A NETWORK

Network Characteristics

Consortia in the Network of Investigator Networks are comprised of between 5 and 521 teams. Subject numbers range from 3,000 to over half a million. Elements deemed essential for launching a network are a strong scientific rationale, the agreement of all teams to work together and combine data on overarching research questions, and the ability to support initial communication, coordination, identification, and recruitment of partners. True integration of disciplines can be challenging because different disciplines are typically housed in discrete departments and have different scientific cultures. Interdisciplinary training is important for bridging these gaps.

Established networks have coalesced through different processes. Frequently, the initiation of a network includes the gathering of information on available resources from several groups of investigators actively involved in research in the same field. Dissemination of information on integrated research aims, resources, and possible contributors ultimately leads to the identification of specific projects to be pursued. This process creates a forum for scientific exchange and more targeted collaborations.²⁸ Networks tend to expand their membership over time and loss of partner teams is uncommon.^{29,30}

TABLE 1. Challenges Faced by Networks of Investigators in Human Genome Epidemiology and Possible Solutions

Major Challenges	Possible Solutions
Resources for establishing the initial infrastructure, supporting consortia implementation, and adding new partners	New and more flexible funding mechanisms: planning grants, collaborative research grants Coordination among national and international funding agencies and foundations Appropriate evaluation criteria for continuation of funding
Coordination: minimize administration to maximize scientific progress and avoid conflicts	Clear leadership structure: steering committee and working groups Early development of policies and processes Cutting-edge communication technology
Selection of target projects	Questions that can be uniquely addressed by collaborative groups Preliminary supportive evidence High-profile controversial hypothesis Biologic plausibility Genomewide evidence
Variable data and biospecimen quality from participating teams	Eligibility criteria based on sample size Sound and appropriate study design Accurate phenotype outcome and genotype assessments State-of-the-art biospecimen repositories
Handling of information from nonparticipating teams and of negative results	Integration of evidence across all teams and networks in a field Comprehensive reporting to maintain transparency Curated updated encyclopedia of knowledge base
Collection, management, and analysis of complex and heterogeneous data sets	Central informatics unit or coordinating center “Think tank” for analytic challenges of retrospective and prospective data sets Centralization of genotyping Standardization or harmonization of phenotypic and genotypic data Standardization of quality control protocols across participating teams Rapid integration of evolving high throughput genomic technologies Consideration of centralized platforms Maximizing use of bioresources Public-private partnerships Development of analytic approaches for large and complex data sets
Anticipating future needs	Web-based communication: websites and portals Teleconferences and meetings support
Communication and coordination	Upfront definition of publication policies Mentorship of young investigators Change in tenure and authorship criteria
Scientific credits and career development	Data-sharing plan and policies Support for release of public data sets Availability and dissemination of both “positive” and “negative” results Encyclopedia of knowledge
Access to the scientific community at large and transparency	Review criteria appropriate for interdisciplinary large science Education of peer scientists to consortia issues Inclusion of interdisciplinary expertise in initial review groups
Peer review	Anticipation of data and biospecimen sharing requirements and careful phrasing of informed consent Sensitivity to local and national legislations
Informed consent	

Although network membership tends to be inclusive, there is concern that inclusion of flawed data jeopardizes the validity of the collaborative results. For this reason, some consortia have eligibility criteria based on appropriateness of study design and phenotypic accuracy.

Organization and Coordinating Centers

Networks use different models of steering and coordination. Working groups focused on specific topics are com-

mon within the largest networks. For example, the International Head and Neck Cancer Epidemiology (INHANCE) network³² requires all members to participate in at least one of 7 working groups that focus on scientific issues or projects such as age at cancer onset, nonsmokers and nondrinkers, tobacco and alcohol, genetics and DNA repair, human papilloma virus prognosis and survival, and occupational factors. The Genetics of Melanoma (GenoMEL) network³³ has a Steering Committee, a Scientific Advisory Board, a Patient

Advocacy Group and an Ethics Committee as well as several topic-specific working groups. Some networks have separate statistical, genetic, and clinical coordinating centers, whereas others centralize these functions. A primary coordinator or chair and a small steering group are usually essential for the network to operate efficiently. Sometimes it is difficult to trace in detail what happens at the local level of participating sites. Minimizing and streamlining administration to maximize the conduct of science is essential.

Funding

Funding sources include governmental and public health agencies as well as private foundations. Funding from for-profit companies and full partnership with industry-sponsored teams has been rare, although some consortia have partnered with private companies for specific projects. For example, the Colon Cancer Family Registry worked with specific companies to perform a systematic mutational analysis of the participants enrolled.³⁴ Funding, especially for infrastructure, is a key limiting factor. Difficulties also exist occasionally for obtaining funding to support activities beyond the originally proposed specific projects despite demonstrated productivity of the network. Some consortia have a single source for primary funding (typically National Institutes of Health or European Commission grants), but most networks have diverse, sometime project-specific, sources of funding. For example, the Birth Cohorts Consortium had a total of 64 funders over the last 8 years. In some countries, participation in a consortium can constitute a strong leverage to obtain national funds.

STANDARDIZATION WITHIN THE NETWORK

Data Management

Efficient and accurate data management is very important because poor-quality data from one or more teams may undermine an otherwise excellent collaboration. Data typically flow to one coordinating center, but some consortia have multiple data coordinating centers with complementary functions.

Networks use various data quality assurance practices and checks for logical errors and inconsistencies. Networks that have invested heavily in quality assurance believe that the effort was worthwhile, because errors may occur even under the best circumstances.³⁵ Logical errors (inconsistencies in the contributed data) are usually easy to identify and readily solved through communication with the team investigators. Examples include out-of-range values, inversion of coding of phenotypes, improper or inconsistent allele calling, and inconsistent crosscoding in databases. Logical errors may reveal deeper problems with contributed data. Queries regarding missing data may yield additional information with some additional effort from the team. Some consortia have instituted in-person training for collecting genotype and phenotype data in addition to ongoing quality control checks. Some networks have developed and published explicit policies of quality assurance for phenotype or genotype data.²⁵

Standardization or Harmonization of Phenotypes and Other Measurements

Data standardization is best implemented at the beginning of a “de novo” collaborative study, when tools for data collection and definition of data items are developed. Data standardization achieves agreement on common data definitions to which all data layers must conform. Each data item is given a common name, definition, and value set or format. When standardization is not possible (eg, different questionnaires or criteria have been used historically by different teams), harmonization of data items is suggested—and sometime required by the funding agencies. Data harmonization is useful when data sets are already collected from originally independent studies focusing on similar questions or field of inquiry. The harmonization process seeks to maximize the comparability of data from 2 or more information systems with the goal of reducing data redundancy and inconsistencies as well as improving the quality and format of data.

Standardization or harmonization is crucial for a network to perform better than single studies, and these processes increase the credibility of the derived evidence. Phenotypes and other nongenetic measurements may be difficult to standardize across teams. For example, Parkinson disease has several sets of accepted diagnostic criteria and teams may use different criteria that have high concordance. It is often challenging to reassess phenotype using alternative criteria. In some diseases, there may be no consensus regarding the most important phenotypes to study. For example, 21 pharmacogenetic studies in asthma analyzed 483 different end points.³⁶

Conversely, the assembled data of some networks have been used to define subphenotypes of disease that would not have been evident with lower statistical power.³⁷ Networks may help achieve harmonization, even when single-team studies have been inconsistent in preferred definitions and outcomes. For example, in the HIV consortium, access to primary data allowed for harmonized definitions of seroconverter and seroprevalent subjects and for the outcome (clinical AIDS),³¹ although these variables had been defined inconsistently by the teams. In contrast, the InterLymph consortium standardizes the diagnosis of lymphoma subtypes through a coordinated review of a subset of slides from each numbered study.³⁸ One criterion of the importance and success of a network may be its ability to adopt standards for phenotypes and covariates to prevent the use of inconsistent definitions in subsequent studies.

In some networks, phenotypes are assessed in prospectively ascertained cases or through an extensive reexamination of phenotypes of existing cases. Consortia also use training sessions on phenotyping, photographs (eg, for moles in melanoma family members), and central review to enhance consistency of data.

Standardization of Genotypes

Most networks have not performed central genotyping of all samples, but exceptions exist.^{32,39} Shipping specimens is sometimes challenging in collaborations among geographically dispersed teams and regulatory considerations may also

prohibit centralized genotyping. For example, some teams are prohibited from shipping specimens by their protocol, local legislation, or their funding agency. Several networks use a semicentralized approach in which some teams ship their samples to a central laboratory, whereas others perform onsite genotyping.

Quality control of genotype results is usually straightforward, but additional checks are required in a multiteam collaboration. Some networks use published genotype data without quality checks beyond what each individual team implemented in their laboratory (eg, repeat genotyping of a random sample of specimens). In the absence of centralized quality control, consortia must depend on post hoc analyses such as deviation from Hardy-Weinberg equilibrium proportions in the controls,⁴⁰ to identify possible genotyping (or other) errors. Large between-study heterogeneity in the final analyses may also reflect measurement errors. However, sizeable errors may still be missed with these methods.

Several networks, including the Public Population Project in Genomics (P3G), check genotype results through exchange of blinded samples between groups. Another approach is to ship samples of known (ideally sequence-verified) genotypes to all participating laboratories. Alternatively, a sample of specimens that were genotyped locally may be shipped to a central laboratory for confirmation. Experience suggests that the reliability of each laboratory should not be taken for granted. Serious errors have occurred (eg, inverse reporting of genotype results that produces an inverse association) that could only be detected by rigorous quality control mechanisms. Error rates may be considerable even for single nucleotide polymorphisms and can depend on a laboratory's methodology and expertise. This is particularly relevant because most gene-disease associations have modest effect sizes that could be obscured by small laboratory errors.

OTHER ORGANIZATIONAL ISSUES

Communication and Web Site Development

Networks use face-to-face meetings, e-mail, teleconferences, and password-protected web sites to communicate with an increasing preference for electronic communication (for details, see web sites¹⁴). Web sites promote visibility and diffuse basic information on the network, activities, and products (eg, publications). Portals provide password-protected access for more sensitive information, which is essential to communication within and between teams as well as venues for private scientific interaction with fellow members. Some networks have developed principally as registers of data from multiple groups and their data management is entirely web-based such as the meta-analysis on DNA repair and cancer risk.⁴¹

Publication and Authorship

Explicit review and publication policies are best established early in the life of a network to avoid later dissent. For each manuscript, a core writing team is essential for developing an initial draft and incorporating comments from co-authors. Most consortia use individual-name authorships, which result in a long list of authors. The first author is

typically the leader of the specific project. Some networks use tiered authorship (authors and separate lists of additional contributors and separate acknowledgments). Group authorship may also be used, but errors in tracking publications in PubMed and the Science Citation Index may occur.⁴² Intellectual property rights may also be an issue in consortia. A carefully crafted agreement involving all partners should be formulated at the outset.

Authorship position and principal investigator status on funded grants are critical for promotion of junior investigators. In the long run, networks will likely produce fertile ground for career development by assuring expert interdisciplinary mentorship and providing opportunities for developing productive scientific collaborations, but in emerging consortia, more senior investigators tend to assume major responsibilities and receive the corresponding authorship credit and grant funding. Some consortia have developed explicit policies of ensuring opportunities for young investigators. Changes in funding mechanisms, tenure criteria, and publication credit are needed to support consortia as a tool for both the rapid advancement of scientific knowledge and the development of new independent investigators.⁴³

Access to Data and Nonselective Availability of Data

Network-developed data and resources should be accessible to the larger scientific community and networks should develop data-sharing policies that support this requirement. Standardization of data-sharing policies is needed and could be facilitated by regulations and policies formulated by funding agencies.⁴⁴

It is important that both "positive" and "negative" results be reported to avoid publication bias.⁴⁵ By their very nature, networks may be the last line of defense against selective reporting and resulting publication biases and should strive to identify and include high-quality, but previously unpublished, data.

Peer Review Process

Interdisciplinary science requires interdisciplinary peer review. Education of peer scientists and establishment of initial review groups with appropriate interdisciplinary expertise are vital to evaluate accurately the merit of consortia proposals. Interdisciplinary research teams take time to assemble and require unique resources.⁴⁶⁻⁴⁸ Targeted funding mechanisms may be needed, especially to build infrastructures for emerging consortia. Criteria for evaluation of productivity by funding agencies should take into account the planning and time to establish the necessary infrastructure.

Informed Consent

Networks need flexibility to address emerging scientific questions. Informed consent should allow data sharing and support broad areas of research conducted by multiple investigators at different institutions in different countries. Examples of elements to be included in such informed consent have been published and adopted by some existing consortia.⁴⁹ However, the variable requirements of Institutional Review Boards at different institutions in considering the incorpora-

tion of these elements and the great heterogeneity of privacy legislation at the state, national, and international level may complicate data and biospecimen sharing in large consortia.⁵⁰

OTHER CHALLENGES AND OPPORTUNITIES

The meeting participants identified a number of additional challenges. For example, inclusiveness criteria are challenging and should be balanced against proper quality assurance. Single teams should be free to pursue their research priorities, and their promising results may then be replicated by the consortium at large. All “negative” results should be fully recorded, preferably in an open access environment, to avoid wasted duplication of effort and confusion in the field. Plurality may also reflect the existence of multiple networks in the same field with similar or very different designs. Accurate registration of membership may mitigate overlap and maximize comparison and replication of results. Upfront study registration has been adopted for clinical trials: ClinicalTrials.gov accepts nonrandomized studies and already has 4,000 or more in its database. Central tracking of genomewide association studies is being planned by the National Institutes of Health as a means to minimize publication and reporting biases, maximize transparency and data access rapidly advance research, and maximize funding allocation.²² Rapid and continuous integration of cutting-edge genomic and other technologies is a challenge. This may require the adoption of centralized technology platforms, which may be supported by public–private partnerships such as the GAIN initiative.⁴⁶ Long-term planning should take into account the fact that laboratory techniques are rapidly becoming cheaper and easier to apply on a large-scale basis. The development, maintenance, and standardization across teams of high-quality biologic repositories (or “biobanks”) are a further challenge. The ultimate goal is to maximize bioresources through various valid strategies such as immortalized cell lines, whole genome amplification, pooling, tissue microdissection, or multiplex microarrays as deemed appropriate.

Many of the challenges facing networks, if properly addressed, may yield opportunities, as summarized in Table 2.

TABLE 2. Potential for Networks to Contribute to Research Progress in Human Genome Epidemiology

Improve the quality of primary studies
Improve the standards of clinical, laboratory, and statistical methods
Strengthen the quality of international collaborative studies, and thereby reduce language and publication biases ⁵⁰
Provide empirical evidence for developing the optimal criteria for grading the credibility of evidence for genetic association studies ⁵¹
Facilitate testing of between-studies heterogeneity in both allele frequencies and size of genetic effects across participating groups studying different populations
Facilitate replication of complex associations involving entire loci or pathways in large-scale data sets
Support methodologic development

CONCLUSIONS

The HuGENet Network of Investigators Networks seeks to provide an open forum for communication and sharing of expertise in statistical and laboratory methods, policies, and procedures among consortia. Consortia are encouraged to create a core registry that would include basic information on their participating teams and on the characteristics of their studies and target populations. This wider knowledge base would improve efficiency in planning further studies and allow for faster replication of results needing validation. Another HuGENet Network of Investigator Networks effort aims at developing an online encyclopedia of genomic epidemiology, maintaining updated information on results from ongoing studies. Such “synopses” of evidence are underway for several diseases, experimenting with various formats that would be comprehensive and flexible enough to cover the needs of a rapidly developing field.^{52–54} Ultimately, if interdisciplinary “large science” human genome epidemiology is to succeed, academic institutions, funding agencies, and scientific journals must incorporate policies, processes, and rewards that support team science while respecting individual creativity. This will require a fundamental change, which is already afoot, from a research culture of “rugged individualism” to one of team work.

ACKNOWLEDGMENTS

From the *Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD; the †Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA; the ‡Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; the §National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; the ||Canada Research Chair in Human Genome Epidemiology, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada; the ¶MRC Biostatistics Unit, University of Cambridge, Cambridge, U.K.; the #Public Health Genetics Unit, Strangeways Research Laboratory, Cambridge, U.K.; the **Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY; the ††International Agency for Research on Cancer, Lyon, France; the ‡‡Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, TX; the §§Center for Human Genetics, Institute of Molecular Medicine and School of Public Health, University of Texas, Houston, TX; ||||Renal Research Laboratories, Manchester Institute of Nephrology and Transplantation, Royal Infirmary, Manchester, U.K.; the ¶¶University of California, Berkeley, CA; the ##Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK; the ***Department of Public Health and Primary Care, University of Cambridge, Cambridge, U.K.; the †††Department of Social Medicine, University of Bristol, Bristol, U.K.; ‡‡‡Albert Einstein College of Medicine, Bronx, NY; the §§§World Health Organization, Geneva, Switzerland; the |||||Department of Dermatology, Leiden University Medical Center, Leiden, The Netherlands; ¶¶¶Harvard School of Public Health, Boston, MA; the

###Department of Epidemiology and Public Health, Imperial College, London, U.K.; the ****Department of Public Health Science and General Practice, University of Oulu, Oulu, Finland; the ††††Department of Radiation Sciences, Oncology, Umea University Hospital, Umea, Sweden; the ‡‡‡‡Department of Neurology, Mayo Clinic, Rochester, MN; the §§§§Genetic Epidemiology Division, CR-UK Clinical Centre, Leeds, U.K.; the |||||University of Pittsburgh Medical Center, Pittsburgh, PA; the ¶¶¶¶Departments of Internal Medicine and Epidemiology & Biostatistics, Erasmus MC, Rotterdam, the Netherlands; the #####ISI Foundation, Torino, Italy; the *****Medical Research Council Epidemiology Unit, Elsie Widdowson Laboratories, Cambridge, U.K.; the †††††Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece; and the ‡‡‡‡‡Department of Medicine, Tufts University School of Medicine, Boston, MA.

REFERENCES

1. Large Scale Biomedical Science: Exploring Strategies for Future Research; Committee on Large Scale Science and Cancer Research. IOM Report; 2003.
2. Seminara D, Ostram GI. Genetic epidemiology of cancer: a multidisciplinary approach. *Genet Epidemiol.* 1994;11:235–254.
3. Relationship of blood pressure, serum cholesterol, smoking habit, relative weight and ECG abnormalities to incidence of major coronary events: final report of the pooling project. The Pooling Project Research Group. *J Chronic Dis.* 1978;4:201–306.
4. Kreeger K. Consortia 'big science': part of a paradigm shift for genetic epidemiology. *J Natl Cancer Inst.* 2003;95:640–641.
5. Khoury MJ. The case for a global human genome epidemiology initiative. *Nat Genet.* 2004;36:1027–1028.
6. Collins FS. The case for a US prospective cohort study of genes and environment. *Nature.* 2004;429:475–477.
7. Ioannidis JPA, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol.* In press.
8. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6:95–108.
9. Thomas DC, Haile RW, Duggan D. Recent developments in genome-wide association scans: a workshop summary and review. *Am J Hum Genet.* 2005;77:337–345.
10. Caporaso NE. Why have we failed to find the low penetrance genetic constituents of common cancers? *Cancer Epidemiol Biomarkers Prev.* 2002;11:1544–1549.
11. Timpson NJ, Lawlor DA, Harbord RM, et al. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet.* 2005;366:1954–1959.
12. Rogers S, Dowling E, Valle C, et al. *The Trends and Development in Consortia as a Tool for Genetic Research in Epidemiology.* Proceedings of AACR Frontiers in Cancer Prevention and Research; 2005;C79.
13. Ioannidis JPA, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. *Am J Epidemiol.* 2005;162:302–304.
14. Human Genome Epidemiology web sites. Available at: www.hugenet.org.uk/networks/list.html (Cambridge) and www.dhe.med.uoi.gr/hugenet.htm (Ioannina). Accessed July 10, 2006.
15. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst.* 2004;96:434–442.
16. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet.* 2003;361:865–872.
17. Ioannidis JPA. Why most published research findings are false. *PLOS Medicine.* 2005;2:e124.
18. Taioli E. International collaborative study on genetic susceptibility to environmental carcinogens. *Cancer Epidemiol Biomarkers Prev.* 1999;8:727–728.
19. Bernstein JL, Teraoka S, Haile RW, et al. WECARE Study Collaborative Group Designing and implementing quality control for multi-center screening of mutations in the ATM gene among women with breast cancer. *Hum Mutat.* 2003;21:542–550.
20. Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet.* 2005;77:685–693.
21. Hunter DJ, Riboli E, Haiman CA, et al. National Cancer Institute Breast and Prostate Cancer Cohort Consortium. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer.* 2005;5:977–985.
22. NIH Guide for Grants and Contracts web site. Available at: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-06-071.html>. Accessed June 30, 2006.
23. Thomas DC. Are we ready for genome-wide association studies? *Cancer Epidemiol Biomarkers Prev.* 2006;15:595–598.
24. Raimondi S, Taioli E. APIKIDS: registry of children born after assisted reproductive technologies. *Paediatr Perinatal Epidemiol.* In press.
25. John EM, Hopper JL, Beck JC, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res.* 2004;6:R375–389.
26. NCI web site. Available at: www.cancer.gov/newscenter/pressreleases/cohortconsortium. Accessed June 30, 2006.
27. Ioannidis JP, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. *Nat Genet.* 2006;38:3–5.
28. NCI web site. Available at: <http://epi.grants.cancer.gov/initiatives.html>. Accessed June 30, 2006.
29. GENOMOS web site. Available at: www.genomos.org. Accessed June 30, 2006.
30. Uitterlinden AG, Ralston SH, Brandi ML et al. Large-scale analysis of association between common vitamin D receptor gene variations and osteoporosis: the GENOMOS Study. *Ann Intern Med.* In press.
31. Ioannidis JP, Rosenberg PS, Goedert JJ, et al. International Meta-Analysis of HIV Host Genetics. Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A alleles on HIV-1 disease progression: an international meta-analysis of individual-patient data. *Ann Intern Med.* 2001;135:782–795.
32. IARC web site. Available at: <http://inhance.iarc.fr/>. Accessed June 30, 2006.
33. GENOMEL web site. Available at: <http://www.genomel.org>. Accessed June 30, 2006.
34. CFR web site. Available at: <http://epi.grants.cancer.gov/CFR/>. Accessed June 30, 2006.
35. Pompanon F, Bonin A, Bellemain E, et al. Genotyping errors: causes, consequences and solutions. *Nat Rev Genet.* 2005;6:847–859.
36. Contopoulos-Ioannidis, Alexiou G, Gouvas T, et al. An empirical evaluation of multifarious outcomes in pharmacogenetics: beta2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenetics and Genomics.* In press.
37. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA.* 2005;293:1979–1985.
38. Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. *Lancet Oncol.* 2006;7:27.
39. Andrusis IL, Anton-Culver H, Beck J, et al. Cooperative Family Registry for Breast Cancer studies. Comparison of DNA- and RNA-based methods for detection of truncating BRCA1 mutations. *Hum Mutat.* 2002;20:65–73.
40. Yonan AL, Palmer AA, Gilliam TC. Hardy-Weinberg disequilibrium identified genotyping error of the serotonin transporter (SLC6A4) promoter polymorphism. *Psychiatr Genet.* 2006;16:31–34.
41. <http://130.192.70.30/huge/>.
42. Dickersin K, Scherer R, Suci ES, et al. Problems with indexing and citation of articles with group authorship. *JAMA.* 2002;287:2772–2774.

43. NIH multi-P.I. initiative. Available at: http://grants.nih.gov/grants/multi_pi/. Accessed June 27, 2006.
44. National Institutes for Health. Data Sharing Policy web site, p 17. Available at: http://grants2.nih.gov/grants/policy/data_sharing/. Accessed June 10, 2006.
45. Ioannidis JP. Journals should publish all 'null' results and should sparingly publish 'positive' results. *Cancer Epidemiol Biomarkers Prev.* 2006;15:186.
46. GAIN initiative. Available at: www.fnih.org/GAIN/GAIN_home.shtml. Accessed June 27, 2006.
47. NIH Roadmap interdisciplinary consortia link, p 18. Available at: <http://grants2.nih.gov/grants/guide/notice-files/NOT-RM-05-006.html>. Accessed June 27, 2006.
48. European Community Framework. Available at: www.infoexport.gc.ca/science/eu_7fp-en.htm?jsessionid=IZHW1NF2GMCHICTHNCACFEQ. Accessed June 27, 2006.
49. Daly MB, Offit K, Li F, et al. Participation in the cooperative family registry for breast cancer studies: issues of informed consent. *J Natl Cancer Inst.* 2000;92:452–456.
50. Betancourt D, Dowling E, Seminara D; International Consortium on Prostate Cancer Genetics. Data sharing and informed consent in genetic epidemiology consortia. Poster presentation, 9th International Meeting on the Psychosocial Aspects of Genetic Testing for Hereditary Cancer, 2005:19.
51. Pan Z, Trikalinos TA, Kavvoura FK, et al. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med.* 2005;2:e334.
52. Ioannidis JP. grading the credibility of molecular evidence for complex diseases. *Int J Epidemiol.* 2006;35:572–828.
53. De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med.* 2004;351:1250–1251.
54. Embracing risk. *Nat Genet.* 2006;38:1.