A road map for efficient and reliable human genome epidemiology


Networks of investigators have begun sharing best practices, tools and methods for analysis of associations between genetic variation and common diseases. A Network of Investigator Networks has been set up to drive the process, sponsored by the Human Genome Epidemiology Network. A workshop is planned to develop consensus guidelines for reporting results of genetic association studies. Published literature databases will be integrated, and unpublished data, including ‘negative’ studies, will be captured by online journals and through investigator networks. Systematic reviews will be expanded to include more meta-analyses of individual-level data and prospective meta-analyses. Field synopses will offer regularly updated overviews.

1Clinical 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 45110, Greece. 2Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. 3Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA. 4Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada. 5Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge CB2 2SR, UK. 6Public Health Genetics Unit, Strangeways Research Laboratory, Cambridge CB1 8RN, UK. 7Department of Epidemiology and Community Medicine, University of Cambridge, Cambridge CB2 2SR, UK. 8Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA. 9Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. 10Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada. 11Renal Research Laboratories, Manchester Institute of Nephrology and Transplantation, Royal Infirmary, Manchester M13 9WL, UK. 12University of California, Berkeley, California 94720-7360, USA. 13Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WCIE 7HT, UK. 14Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. 15Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol, BS8 2PR, UK. 16March of Dimes, White Plains, New York 10605, USA. 17World Health Organization, Geneva CH-1211, Switzerland. 18Department of Dermatology, Leiden University Medical Center, 2333 AL Leiden, The Netherlands. 19Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, USA. 20Harvard School of Public Health, Boston, Massachusetts 02115, USA. 21Department of Epidemiology and Public Health, Imperial College, London W2 1PG, UK. 22Department of Public Health Science and General Practice, University of Oulu, Oulu, Finland. 23Department of Radiation Sciences, Oncology, Umea University Hospital, Sweden. 24Department of Neurology, Mayo Clinic, Rochester, Minnesota 55905, USA. 25Genetic Epidemiology Division, CR-UK Clinical Centre, Leeds LS8 7FT, UK. 26Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. 27Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, Maryland 20892, USA. 28University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15232, USA. 29Departments of Internal Medicine and Epidemiology & Biostatistics, Erasmus MC, Rotterdam 3000DR, The Netherlands. 30Institute for Scientific Interchange Foundation, Torino, Italy. 31Medical Research Council Epidemiology Unit, Elsie Widdowson Laboratories, Cambridge CB1 9NL, UK.

e-mail: jioannid@cc.uoi.gr and muk1@cdc.gov
Genetic epidemiologists and geneticists face the challenge of creating an efficient and reliable compilation of the evidence for genetic risk contributions to common human diseases (‘a risk engine’). Although data relating DNA sequence variation to disease states and/or intermediate traits are accumulating exponentially, the current situation is plagued with problems. These include the prevalence of small, underpowered studies, often with flawed designs, suboptimal conduct and biased analyses; selective reporting of ‘positive’ results; lack of standardization among studies; poor reporting of results even from well-conducted studies; and difficulties in assessing interactions with environmental risk factors. Consequently, the research evidence is fragmented, and the interface between epidemiological and other biological evidence is poorly developed. It remains unclear how to keep track of the rapidly evolving evidence across fields that can be defined by disease, genes or exposures, and how to rate the credibility of this evidence.

The Human Genome Epidemiology Network (HuGENet), a global initiative committed to the development and integration of the knowledge base on human genetic variants and health (http://www.cdc.gov/genomics/hugenet), proposes a plan for developing this knowledge base on human genetic variants and health (http://www.cdc.gov/genomics/hugenet). A HuGE-STROBE guideline will offer an objective checklist that can be used for investigators, peer reviewers and editors on improving the quality of the studies and will aid their assessment before publication. Meeting the reporting requirements may result in longer manuscripts, but all statistical steps, assumptions and processing can be documented in supplementary files.

HuGENet has already created a database (HuGEPubLit) that aims to capture published genetic association articles as they are indexed in Medline. As of November 2005, the database has over 18,500 entries. This effort will be extended to other databases such as EMBASE and will seek synergy with ongoing, similar initiatives such as the US National Institutes of Health (NIH)-sponsored Genetic Associations Database (GAD, http://geneticassociationdb.nih.gov). Similar efforts are ongoing in pharmacogenetics and pharmacogenomics (http://www.pharmgkb.org).

These databases need search tools, and they should encompass both published and deposited data (Table 1, Step 3). Retrieving unpublished data is currently very problematic, and unpublished reports often present ‘negative’ results from well-conducted studies. A research environment that promotes and rewards only results that reach formal statistical significance is likely to foster data dredging and will create a distorted literature with very low credibility. Comparisons of primary outcomes defined in trial protocols with those defined in published articles have provided empirical evidence of selective reporting even for randomized controlled trials. Selective reporting of extensive exploratory analyses would be almost impossible to detect in studies of gene-disease associations and related interactions, even by the most sophisticated methodologists and expert peer reviewers. The protocols of these studies are rarely available for scrutiny, and there is currently no formal way for registering analyses in advance of publication.

HuGENet is working with journals and collaborating in efforts to create online journals to encourage publication of ‘negative’ results after appropriate methodological appraisal and with due credit to the investigators. The citable, peer-reviewed Molecule Pages of the Alliance for Cell Signalling...
1. Develop a Network of Investigator Networks
2. Improve study conduct, reporting and harmonization across studies
3. Capture published and unpublished data regardless of ‘positive’ or ‘negative’ results
4. Improve data synthesis methods and integrate the evidence on specific associations
5. Capture and appraise the evidence on the evolving ‘big picture’ across whole fields

Steps | Action items
--- | ---
1. **Develop a Network of Investigator Networks**<br>Create capacity to accomplish steps 2–5 below
2. **Improve study conduct, reporting and harmonization across studies**<br>Share among networks best practices, tools and analytic methods; develop STROBE criteria for genetic association studies (consensus workshop planned for 2006); single studies performed with eventual meta-analysis in mind
3. **Capture published and unpublished data regardless of ‘positive’ or ‘negative’ results**<br>Integrate published literature databases; capture unpublished data (online journals, networks); enhance transparency of methods and critical appraisal; develop comprehensive search engines
4. **Improve data synthesis methods and integrate the evidence on specific associations**<br>Finalize HuGE handbook for conducting systematic reviews; promote methods for meta-analyses of individual-level data; expand database of HuGE reviews and other systematic reviews and meta-analyses; facilitate meta-analyses from consortia
5. **Capture and appraise the evidence on the evolving ‘big picture’ across whole fields**<br>Develop widely accepted criteria for appraising evidence; initiate pilot phase for specific fields; hold consensus meeting for guidelines on grading the evidence in 2006; complete empirical research; publish regularly updated synopses of the knowledge base; identify knowledge gaps

---