



# Can Lessons Learned from Genome-Wide Research be Applied to Nutrition-Wide and Exposure-Wide Evidence?

JOHN P.A. IOANNIDIS

Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and the Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts

Human genomics and human nutrition are two very large domains of epidemiological investigation, but their overlap has remained relatively limited to date. Most studies that address genetic risk factors ignore nutritional and other related exposures, and vice versa. Both domains have not only a history of exciting discoveries, but also a tenuous, if not poor, replication record (Ioannidis, 2005a; Ioannidis et al., 2001). For nutritional and lifestyle epidemiology, randomized trials have failed to confirm a large number of associations proposed by observational data (Ioannidis, 2005b). For genetic epidemiology, a major transformation in the last few years has suggested that most candidate gene associations proposed in the past were likely false-positives (Manolio et al., 2008; McCarthy et al., 2008).

The new transformation of genetic epidemiology has led to a new mode of discovery and validation of associations based on shear massive testing (large-scale studies on large-scale massive-testing platforms) (McCarthy et al., 2008). The list of successfully discovered gene variants that regulate the risk for common diseases continues to grow on a weekly basis (Manolio et al., 2008). It is worthwhile to examine whether some of the lessons of this recent transformation can also be extended to human nutritional and lifestyle epidemiology. There are clear differences between the genomics and nutrition fields. A major difference is that massive measurement platforms are not yet available for nutritional and lifestyle exposures, at least on the large scale available for genomic markers. Nevertheless massive-scale biochemical measurements and massive collection of electronic information (e.g., through mobile telephones or the Internet) may get closer to achieving the high-throughput paradigm that is currently prevalent in genomics.

The correlation pattern between nutritional/lifestyle variables is far denser than the correlation pattern of genomic markers; however, disequilibrium is also considerable in some areas of the genome linkage and poses similar problems in the attribution of causality. In general, markers that arise from genome-wide association studies are only correlates of risk, and may be far from the real causative culprit.

Human genome epidemiology has also made major progress by being able to cut down dramatically in measurement error by imposing very strict quality-control standards. The use of rigorous standards also should be feasible in other areas of epidemiological investigation.

The need for longitudinal measurements and the unavoidable missing information may be different from genomic epidemiology; however, genomic markers also need to be further validated in longitudinal studies, rather than only the case-control designs that have been relatively successful for genetic epidemiology so far (Wellcome Trust Case Control Consortium, 2007). Firm documentation of gene-environment interactions will require examination in very large cohorts and biobanks (Elliott et al., 2008). Thus, a new paradigm of extremely large cohorts and coalitions thereof is due to appear. Repeated measurements are also essential for nutritional epidemiology, as opposed to the genotypic information that is standard and fixed at birth.

Genomic epidemiology has been more eager to adopt explicit consideration of multiplicity issues, which has led to a more uniform adoption of rigorous standards for what is considered appropriately replicated and credible as an association. Similar improvements must also be made in nutritional epidemiology in which traditional significance levels without any correction are

entrenched in the literature. Most associations that are claimed to be significant in traditional epidemiology would have very modest, inconsequential Bayes factors (Ioannidis, 2008) if viewed from a Bayesian perspective and their credibility would likely be very tenuous.

Nutritional and lifestyle epidemiology may also learn useful lessons from the advent and successful application of large international consortia in genomic epidemiology (Seminara et al., 2007) and from the public deposition of data and lack of selective reporting in large-scale genomic databases (Ioannidis, 2007).

In summary, it is unlikely that either genes or exposures alone will be able to reveal much about the risk of human diseases. Integrating information on both sides may be essential to making genuine progress. Very large studies that explicitly measure genetic exposures, nongenetic exposures, and outcomes on a massive scale may be a way to make progress, but they present several challenges to be surmounted.

## REFERENCES

- Elliott, P., Peakman, T. C., and the UK Biobank. (2008). The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int. J. Epidemiol.* **37**: 234–244.
- Ioannidis, J. P. (2005a). Why most published research findings are false. *PLoS Med.* **2**: e124.
- Ioannidis, J. P. (2005b). Contradicted and initially stronger effects in highly cited clinical research. *JAMA.* **294**: 218–228.
- Ioannidis, J. P. (2007). Molecular evidence-based medicine: evolution and integration of information in the genomic era. *Eur. J. Clin. Invest.* **37**: 340–349.
- Ioannidis, J. P. (2008). Effect of formal statistical significance on the credibility of observational associations. *Am. J. Epidemiol.* **168**: 374–383; discussion 384–390.
- Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nat. Genet.* **29**: 306–309.
- Manolio, T. A., Brooks, L. D., and Collins, F. S. (2008). A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118**: 1590–1605.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**: 356–369.
- Seminara, D., Khoury, M. J., O'Brien, T. R., Manolio, T., Gwinn, M. L., Little, J., Higgins, J. P., Bernstein, J. L., Boffetta, P., Bondy, M., Bray, M. S., Brenchley, P. E., Buffler, P. A., Casas, J. P., Chokkalingam, A. P., Danesh, J., Davey Smith, G., Dolan, S., Duncan, R., Gruis, N. A., Hashibe, M., Hunter, D., Jarvelin, M. R., Malmer, B., Maraganore, D. M., Newton-Bishop, J. A., Riboli, E., Salanti, G., Taioli, E., Timpson, N., Uitterlinden, A. G., Vineis, P., Wareham, N., Winn, D. M., Zimmern, R., and Ioannidis, J. P. (2007). Human Genome Epidemiology Network; the Network of Investigator Networks. The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology.* **18**: 1–8.
- Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* **447**: 661–678.