

- 3 Mardis ER, Wilson RK. Cancer genome sequencing: a review. *Hum Mol Genet* 2009;18:163–8.
- 4 Deisboeck TS. Personalizing medicine: a systems biology perspective. *Mol Syst Biol* 2009;5:249.
- 5 Hood L. A doctor's vision of the future of medicine. *Newsweek* 13 July 2009.

Tobias Schulte in den Bäumen
 Institute for Public Health Genomics
 (IPHG), School for Public Health and
 Primary Care (CAPHRI), Maastricht
 University, The Netherlands,
 Muin J Khoury, Director, Office of Public
 Health Genomics, CDC, Atlanta, USA

Correspondence: Tobias Schulte in
 den Bäumen,
 e-mail: t.sidb@maastrichtuniversity.nl

doi:10.1093/eurpub/ckq201

Public health genomics and the challenges for epidemiology

While chronic diseases (e.g. obesity, diabetes, coronary heart disease, cancer, etc), also named 'complex diseases', are caused by the combined effects of multiple environmental and genetic factors, it is estimated that the environmental component plays a major role. Nevertheless, the smaller role of inherited genetic factors may partly be explained by the difficulties in exploring this component until recently. The huge amount of genomics data produced by the fast-developing biotechnologies at an unprecedented speed will probably help in dissecting the genetics underlying these diseases.

During the past 5 years, genome-wide association studies (GWAS) identified hundreds of genomic loci robustly associated with common chronic diseases.¹ While this strategy has provided key novel insight into disease biology within a very short time scale, it will take many years before the impact of GWAS findings can be precisely estimated. GWAS, as currently conducted, is not able to identify rare variants, structural variants (e.g. copy number variation), and loci displaying high level of allelic heterogeneity across populations because current standards require associations to be replicated in independent populations. Hence, there is still a large fraction of the trait heritability to be explained, the so-called 'missing heritability'. The current application of massive parallel sequencing (MPS) is identifying rare variants associated with common chronic diseases. The huge heterogeneity, almost at an individual level, of the genetic alterations found through these new technologies emphasizes the notion of individualized diseases. Epidemiologists are now moving the focus of their analyses from single genetic variants to entire genes and/or pathways.

Thinking on the biological meaning of the genetic variants, there is probably a continuum between rare monogenic diseases with few loci exerting large

effects and common complex traits with a large number of loci with tiny effect sizes. Thus, most of the identified genetic variants have very small effect sizes and probably interact with other genetic variants and with environmental factors, though examples of robust and validated interactions are still rare.¹ Hence, most genetic variants identified to date, when taken individually, are neither necessary nor sufficient to cause disease but they may have a direct effect on intermediate outcomes such as gene expression and/or protein function. Accordingly, epidemiologists need to deal with the design of large studies aiming at integrating genetics, genomics, transcriptomics and proteomics data at the individual level, together with extensive information on environmental factors and behaviours.

Regarding the latter, epidemiologists face the important challenge of assessing the complexity of highly correlated environmental exposures. We do not have platforms able to assess environmental exposures with the same low measurement error as 'omics' (in particular genomics) platforms do. Rather, we continue asking individuals about their lifespan exposures through questionnaires, which represent soft data. Yet, selected biomarkers represent excellent tools to measure environmental exposures with higher accuracy. A change in paradigm is needed, moving from a candidate to an agnostic/exploratory exposure analysis. Incorporating epigenomics (i.e. modifications in DNA methylation of CpG islands, histone acetylation, etc) and metabolomics markers should help in better dissecting the still 'missing exposurome' for most chronic diseases. Tools for standardized collection across centres and across countries will be needed to this end.

Furthermore, the notion that nothing is static during an individual's lifespan is becoming more and more important. The changes along the time apply to environmental exposures and to 'omics'

data. Taking such lifelong modifications into account in epidemiological studies is going to be a very difficult task that will necessitate to closely monitor individuals. Once the data is available, its modelling will represent another challenge.

Following the current concept of epidemiological study design and statistical power requirements, very large sample sizes are needed to explore the underlying biological complexity in a meaningful manner. Being provocative, we could argue that instead of conducting large scale epidemiological studies, epidemiologist should focus on fewer extremely very well characterized and bio-monitored individuals. In any case, the integration of several types of data, from environmental exposures to epigenetics, metabolomics and genomics, requires the development of innovative bioinformatics and data reduction techniques. There is still a long way to go.

The extraordinary development of hypothesis-free (agnostic) approaches should not discourage researchers from conducting targeted candidate gene studies. Both approaches should be viewed as complementary and synergistic. Similarly, although most GWAS have included unrelated people, family-based studies may bring valuable information on transgenerational effects, shared environmental factors and parent-of-origin effects. An example illustrating future challenges in terms of study design can be found in the field of pharmacogenomics that aims at identifying genetic variants involved in drug response in order to improve drug safety and efficacy, thus minimizing side effects. There are large inter-individual variations in the activity of enzymes involved in drug metabolism and transport that are in large part genetically determined. In contrast to the small effect of the identified variants associated with the risk of common complex traits, the effect of

pharmacogenomic-related variants may be larger and clinically relevant. The recently launched Clarification of Optimal Anticoagulation through Genetics (COAG) double-blind, randomized controlled trial² will ascertain whether adapting the dose of warfarin therapy based on genetic variants located within the *CYP2C9* and *VKORC1* genes may improve patient care as compared with a clinically-guided dosing algorithm. Designing such a trial is particularly challenging because the power to detect a pre-specified between-group difference will depend on the genetic makeup of the participants. The challenge comes from the fact that allele frequencies may vary substantially across ethnic groups. To recommend genetic testing, investigators will need to demonstrate that drug dosing based on genetic information significantly reduces costs and morbidity.³

Focusing on a single disease or on a single trait does not allow understanding the full range of phenotypes associated with many genes, for which pleiotropic effects have been described (i.e. one gene may be involved in both cancer and cardiovascular disease). Hence, an additional challenge for epidemiologists is collecting extensive phenotypic data, not only at a single point in time, but longitudinally, again. The collection of high quality phenotypes and more comprehensive phenomes are therefore of utmost importance and will be key to better account for the underlying biological complexity of human organisms living in selected environmental conditions.^{4,5} The digitalization of patient's records and imaging technologies, as well as web-based testing, should allow accumulating and linking massive amounts of information for each person. The availability of entire genomes and phenomes may revolutionize the way we classify diseases. There is little doubt that data-gathering technology has dramatically changed and will

continue to largely influence the way epidemiologists conduct research. Making best use of all the available information, without harming study participants (i.e. discrimination by insurance companies or employers, undue access to the data by third parties, etc), will be a challenging task in the years to come.

In conclusion, recent advances in genomics have highlighted the polygenic nature of most common disorders. The effects of these genetic variants also need to be studied taking into account time-dependent environmental and behavioural factors.⁵ As a consequence, any single genetic variant has little impact in terms of disease risk prediction. Yet, polygenic risk scores in relation to continuous traits (i.e. BMI, blood lipid, blood pressure, etc) should stimulate public health researchers to change paradigms and consider integration of multilayer biological data, dynamic designs, agnostic approaches, as well as using quantitative measures in assessing both exposures and outcomes (i.e. continuum of affectedness). Such quantitative thinking leads to a public health model that focuses on prevention on a continuous scale rather than just treating cases. Looking at multivariate continuous dimensions rather than clinical diagnoses using arbitrary cut-offs represent more powerful approaches to decipher the complex etiological mechanisms leading to human diseases.

Large inter-disciplinary teams are needed to properly design studies and collect, store and analyse high-throughput data. Whereas the prices of 'omics' data production have dramatically come down, the costs of data storage and analysis are very high and often tend to be underestimated. Unless studies are not well funded, epidemiology will not be able to assume the challenges mentioned above. High-quality and continuously updated education programmes are needed to ensure that researchers and health-care

professionals are able to critically appraise research findings in the 'omics' fields, including 'epi-omics'.

Funding

Murielle Bochud is supported by the Swiss School of Public Health Plus, Núria Malats by EU-FP7-PHGEN II #2008302.

References

- 1 Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166–76.
- 2 French B, Joo J, Geller NL, et al. Anticoagulation Through Genetics Investigators CC. Statistical design of personalized medicine interventions: the Clarification of Optimal Anticoagulation through Genetics (COAG) trial. *Trials* 2010;11:108.
- 3 Hall WD, Mathews R, Morley KI. Being more realistic about the public health impact of genomic medicine. *PLoS Med* 2010;7:e1000347.
- 4 Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet* 2010;11:855–66.
- 5 Ioannidis JP, Loy EY, Poulton R, et al. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med* 2009;1:7ps8.

Murielle Bochud¹, Núria Malats²

¹Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland, ²Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain

Correspondence: Núria Malats, Grupo de Epidemiología Genética y Molecular, Programa de Genética del Cáncer Humano, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro, 3, 28029-Madrid, España. e-mail: nuria@cnio.es

doi:10.1093/eurpub/ckq203

A European view on the future of personalised medicine in the EU

At the European Health Forum Gastein in October 2010, personalized medicine was presented as a new paradigm for health care in Europe. I had the opportunity to participate in this session and to listen to the views of my esteemed co-participants on what can be done at the European level to achieve implementation of personalized medicine

approaches and other emerging technologies.

Personalized medicine is becoming an ever more important area for the application of genome-based information and technologies identified under public health genomics. The scope and vision of public health genomics, according to the 2005 Bellagio

Conference¹ is 'the responsible and effective translation of genome-based knowledge and technologies for the benefit of population health'.

This concept encompasses both population and individual health. It covers prevention, treatment and care.

Genomics is clearly an area of innovative research that seeks to close the