Eur J Epidemiol (2008) 23:707–710
DOI 10.1007/s10654-008-9293-8

COMMENTARY

Is the time right for translation research in genomics?

A. Cecile J. W. Janssens

Published online: 21 October 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Genome-wide association studies are rapidly unraveling genetic susceptibility variants that are implicated in the etiology of common multifactorial diseases such as coronary heart disease, type 2 diabetes, non-familial forms of breast cancer and age-related macular degeneration [1]. Expectations about the future impact of these discoveries on preventive and clinical health care practice are high [2, 3]. Future use of genetic tests is foreseen for the prediction of disease susceptibility, targeting pharmacotherapy and tailoring lifestyle and health behavior recommendations. Fueled by the enormous progress in gene discovery, many researchers are already investigating the prediction of common diseases based on *genetic profiling*, the simultaneous testing of multiple susceptibility variants [4], and an increasing number of companies already offer personalized lifestyle health recommendations and nutritional supplements based on clients' genetic profiles [5]. Despite the current euphoria, the predictive value of genetic profiling is still limited for most disorders, with only some promising exceptions. [4, 6–9] The major limitation to date is that only a fraction of the genetic factors involved have been identified, for most disorders less than 20 [1], explaining not more than a few percentages of the heritability.

While we may expect that a large number of genetic variants will be discovered in the next few years, establishing a solid evidence base for genomics applications in clinical and public health care may take longer given the number of steps to be taken. Khoury and colleagues have described a framework for the continuum of translation research that is required to move genomics research findings to clinical and public health applications that benefit population health [10]. The four phases of translation researches include (1) translation of basic genomics research into a potential health care application; (2) evaluation of the application for the development of evidence-based guidelines; (3) evaluation of the implementation and use of the application in health care practice; and (4) evaluation of the achieved population health impact [10].

Translation research in genomics starts after gene discovery [10]. In common diseases, where numerous genetic factors may be implicated, genes are discovered by demonstrating robust genetic association, not in a single study but in meta-analyses or pooled analyses of large-scale studies [11–13]. A major challenge in common diseases is to decide when we have discovered sufficient genetic variants to begin translation research. One may argue that now the time is right because the studies so far likely have identified the common variants with the strongest effects and that further studies will only add weak susceptibility variants. For instance the complement factor H gene was the first common gene discovered to be involved in age-related macular degeneration (AMD) using not more than 100 patients and 50 controls [14]. Typical gene discovery studies include 1,000s of patients and are able to detect variants with odds ratios as low as 1.05–1.10. Yet, also a very large number of weak susceptibility variants may further improve risk prediction [15]. Furthermore, stronger genetic effects may still be found for gene-gene and gene–environment interactions. Many groups of researchers are currently pooling their data in large consortia, which together will have sufficient power to model and detect interactions. Another avenue to pursue is to target more rare variants with strong genetic effects in specific populations. Genetic associations may not only differ between

A. C. J. W. Janssens (✉)
Department of Epidemiology, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands
e-mail: a.janssens@erasmusmc.nl

ethnic groups, but also within. In Europe, there may be differences in genetic and other risk factors between northern and southern countries (e.g., multiple sclerosis, breast and other cancers, lipids), which asks for gene-disease association studies in specific populations, e.g., [16, 17] also because true genetic heterogeneity between populations may later impact the global applicability of predictive genetic tests. Because gene–gene and gene–environment interactions have not been extensively studied to date, further major advances in unraveling the genetic basis of common diseases may certainly be expected.

All recent studies that investigated the combined predictive value of multiple genetic variants were phase 1 studies. In most studies, the per allele effects of the risk genotypes typically ranged from 1.1 tot 1.4, except for AMD and hypertriglyceridemia [4, 6, 8]. From an epidemiological perspective, investigating the predictive value of a limited number of susceptibility genes with weak effects seems somewhat overoptimistic as a priori high predictive value is not expected [18]. The predictive value of genetic profiling, often investigated in terms of the discriminative accuracy indicated by the area under the receiver operating characteristic curve (AUC), is determined by the number of variants, the frequency of the risk genotypes and their strength of association to disease risk [15]. To reach appreciable predictive value for genetic profiling, we either should be able to include up to tens or hundreds of weak susceptibility genes or a few variants with strong effects as in AMD and hypertriglyceridemia [15]. These variants can be single gene effects, but can also be derived from gene–gene or gene–environment interactions. In the absence of stronger genetic risk factors, phase 1 studies on the predictive value of genetic profiling will continue to yield disappointing results.

Phase 2 research specifies that genetic profiling is evaluated in the population of its intended use. Assessment and replication of the predictive value in independent populations is always important, but particularly when the combined association of multiple variants is initially demonstrated in a case–control study. Most genetic association studies are conducted in case–control studies, and often these include highly selected cases (familial, early onset) and controls (persons with no pathology late in life) to maximize the statistical power. Such studies are more likely to overestimate the combined effect of multiple genes, and extrapolation of estimates from such well-defined study populations to a general population may not be possible [19]. For example, Maller et al. reported that individuals who had risk variants on five variants had 285-fold higher risk of age-related macular degeneration (AMD) then individuals who had none [6]. Yet, they compared two extreme groups, namely those with end-stage AMD and those with no or fewer than 10 small drusen without pigment abnormalities and they did not include patients with early features of AMD [7]. Although this design is powerful and valid for gene discovery, the findings are not informative for the evaluation of genetic testing. Genetic testing for AMD should be evaluated in a prospective cohort study, either a general population cohort of elderly if the intended use is to predict end-stage ÁMD in asymptomatic individuals, or a sub-cohort of patients with early AMD if the intended use is to predict worse prognosis [7]. In contrast to what is common in gene discovery research, evaluations of the predictive value in population-based cohorts do not necessarily need extremely large datasets, as minimal predictive value or minimal improvements in predictive value are generally not of interest from a clinical or public health perspective. Numerous epidemiological cohorts are available and are sufficiently large for this purpose. Examples of population-based cohort studies include the Framingham heart study, European Prospective Investigation into Cancer and Nutrition, the Rotterdam study and LifeLines among adults, and the Avon Longitudinal Study of Parents and Children, Generation R study, and Norwegian Mother and Child Cohort study among mothers and newborns [20–27]. Most of these studies already include extensive genotype data using high throughput genotyping arrays.

Evaluating genetic profiling in the population of its intended use requires that the intended use is already known. The question who will be tested and for what purpose is essential in the evaluation of the usefulness of genetic profiling [28], and becomes particularly relevant in phase 2 research. Genetic profiling may be used for targeting preventive or therapeutic interventions to subgroups, either to individuals who have the highest risk of disease or the worst prognosis or to individuals who benefit most from the intervention. Depending on this intended use, genetic profiling should predict risk of disease/prognosis or treatment response. Phase 2 research can investigate whether effective preventive or therapeutic strategies that are targeted on the basis of traditional risk factors are more effectively and efficiently allocated when risk prediction would be based on genetic factors. Examples include intensive cancer surveillance programs for individuals from high-risk families and breast cancer screening to women over 50 years of age [29]. Phase 2 research is of less interest when the interventions are an obvious benefit for the total population because they target multiple disorders, e.g., smoking cessation and weight reduction, or when no intervention is available. Even when genetic profiling shows high predictive value in phase 2 research, in the absence of effective interventions it will not pass phase 3 and 4.

Notwithstanding the growing availability of commercial genetic testing via the internet, evidence-based

applications of genetic profiling in clinical and public health care practice are still a far future prospect. The recent empirical and modeling studies on the predictive value of genetic profiling have taught us that the identification of stronger genetic associations is paramount for higher predictive value. Most of the scientific attention should, therefore, remain focused on basic genetic epidemiological research, unraveling the genetic basis of common diseases in all its complexity and all its interactions. The progress in genomic research will undoubtedly increase our understanding of disease etiology, leading to the identification of new biomarkers and risk factors that can be used in risk prediction as well as to the development of novel preventive and therapeutic interventions. At this time, phase 1 translation research can contribute by investigating on an aggregate level the conditions to be met for the successful implementation of genetic profiling in public health and clinical practice. For example, to identify the promising applications we need to know when we have sufficient understanding of the contribution of genetic factors to start phase 1 research, and what level of predictive value is considered useful to warrant phase 2 research, e.g., compared to risk prediction based on traditional risk factors, among other questions. The results of such studies can be used to prioritize the translation research agenda so that research time and money are allocated to the most promising of all expected applications.

Acknowledgments The author was sponsored by a fellowship from the Erasmus University Medical Center Rotterdam and by the VIDI grant of the Netherlands Organisation for Scientific Research (NWO).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest. 2008;118:1590–605.
2. Guttmacher AE, Collins FS. Realizing the promise of genomics in biomedical research. JAMA. 2005;294:1399–402.
3. Brand A, Brand H, Schulte In dB. The impact of genetics and genomics on public health. Eur J Hum Genet. 2008;16:5–13.
4. Janssens ACJW, Van Duijn CM. Genome-based prediction of common diseases: advances and prospects. Hum Mol Genet. 2008;17:R166–73.
5. Janssens ACJW, Gwinn M, Bradley LA, Oostra BA, Van Duijn CM, Khoury MJ. A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. Am J Hum Genet. 2008;82:593–9.
6. Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat Genet. 2006;38:1055–9.
7. Despriet DD, Klaver CCM, Van Duijn CM, Janssens ACJW. Predictive value of multiple genetic testing for age-related macular degeneration. Arch Ophthalmol. 2007;125:1270–1.
8. Wang J, Ban MR, Zou GY, Cao H, Lin T, Kennedy BA et al. Polygenic determinants of severe hypertriglyceridemia. Hum Mol Genet. 2008;17:2894–9.
9. Yiannakouris N, Trichopoulou A, Benetou V, Psaltopoulou T, Ordovas JM, Trichopoulos D. A direct assessment of genetic contribution to the incidence of coronary infarct in the general population Greek EPIC cohort. Eur J Epidemiol. 2006;21:859–67.
10. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? Genet Med. 2007;9:665–74.
11. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.
12. Pereira TV, Rudnicki M, Pereira AC, Pombo-de-Oliveira MS, Franco RF. Do polymorphisms of 5,10-methylenetetrahydrofolate reductase (MTHFR) gene affect the risk of childhood acute lymphoblastic leukemia? Eur J Epidemiol. 2006;21:885–6.
13. Zintzaras E, Koufakis T, Ziakas PD, Rodopoulou P, Giannouli S, Voulgarelis M. A meta-analysis of genotypes and haplotypes of methylenetetrahydrofolate reductase gene polymorphisms in acute lymphoblastic leukemia. Eur J Epidemiol. 2006;21:501–10.
14. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308:385–9.
15. Janssens ACJW, Aulchenko YS, Elefante S, Borsboom GJJM, Steyerberg EW, Van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med. 2006;8:395–400.
16. Cottone M, Renda MC, Mattaliano A, Oliva L, Fries W, Criscuoli V, et al. Incidence of Crohn's disease and CARD15 mutation in a small township in Sicily. Eur J Epidemiol. 2006;21:887–92.
17. Portoles O, Sorli JV, Frances F, Coltell O, Gonzalez JI, Saiz C, et al. Effect of genetic variation in the leptin gene promoter and the leptin receptor gene on obesity risk in a population-based case-control study in Spain. Eur J Epidemiol. 2006;21:605–12.
18. Janssens AC, van Duijn CM. Towards predictive genetic testing of complex diseases. Eur J Epidemiol. 2006;21:869–70.
19. Iwasaki M, Yamamoto S, Otani T, Inoue M, Hanaoka T, Sobue T, et al. Generalizability of relative risk estimates from a well-defined population to a general population. Eur J Epidemiol. 2006;21:253–62.
20. Dawber T, Meadors G, Moore F. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health. 1951;41:279–81.
21. Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, Panico S, et al. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. Eur J Epidemiol. 2007;22:129–41.
22. Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, et al. The Rotterdam study: objectives and design update. Eur J Epidemiol. 2007;22:819–29.
23. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, Slaets JP, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol. 2008;23:67–74.

24. Golding J, Pembrey M, Jones R. ALSPAC-The Avon Longitudinal Study of Parents, Children—I. Study methodology. Paediatr Perinat Epidemiol. 2001;15:74–87.
25. Jaddoe VW, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Verhulst FC, et al. The generation R study: design and cohort profile. Eur J Epidemiol. 2006;21:475–84.
26. Jaddoe VW, Bakker R, van Duijn CM, van der Heijden AJ, Lindemans J, Mackenbach JP, et al. The generation R study biobank: a resource for epidemiological studies in children and their parents. Eur J Epidemiol. 2007;22:917–23.
27. Ronningen KS, Paltiel L, Meltzer HM, Nordhagen R, Lie KK, Hovengen R, et al. The biobank of the Norwegian mother and child cohort study: a resource for the next 100 years. Eur J Epidemiol. 2006;21:619–25.
28. Haddow JE, Palomaki GE. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury MJ, Little J, editors. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press; 2003. p. 217–233.
29. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med. 2008;358:2796–803.