

Molecular epidemiology, candidate genes versus genome-wide scans

P. Eline Slagboom · Bastiaan T. Heijmans ·
Marian Beekman · Ingrid Meulenbelt

Published online: 26 September 2007
© Springer-Verlag and NuGO 2007

Keywords Human genetics · Disease ·
Genetic epidemiology · Population studies · Biomarkers ·
Human · Genome scans · Complex disease

Since the completion of the human genome project, there has been a huge effort in establishing maps of genetic markers, especially of the most abundant source of DNA variation, the single nucleotide polymorphism (SNP). A subsequent major quest is to identify the variation in the human genome causally involved in the genetic etiology of complex diseases and traits in concordance with environmental influences such as nutrition. Genetic research of complex traits (diseases, risk factors and intermediate phenotypes) requires large well-characterised study populations of patients and healthy subjects. The aim is to identify at which loci genetic similarity of study subjects corresponds to phenotypic similarity. Genetic similarity can be measured at different genotyping platforms that have different specificities and an ever increasing throughput of genotypes. Studies to explore which genes are involved in a trait include essentially two major approaches. In the hypothesis based approaches the involvement of candidate genes and pathways is being investigated by testing for genetic association in order to find significant frequency differences of marker alleles in cases and controls. Alternatively, explorative approaches

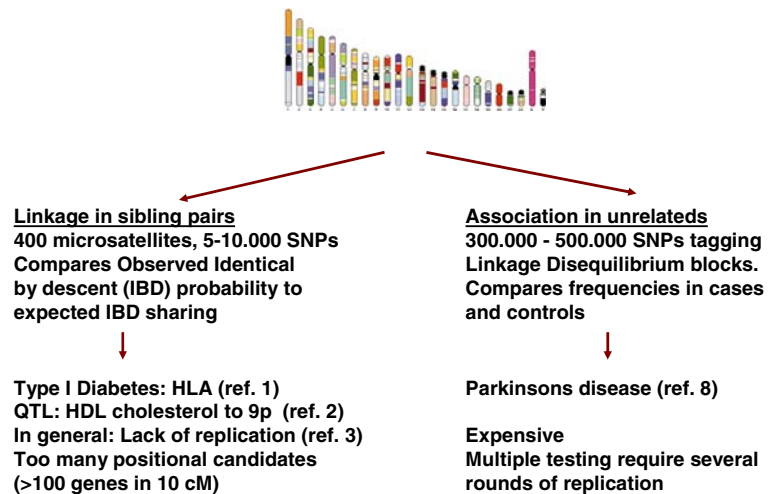
are applied which include genome-wide scans that first localize the gene(s) and subsequently identify the causal genetic variation at that locus. Genome-wide scans can be performed by genetic association analysis or by linkage analysis in related subjects from selected families (Fig. 1).

A decade of family- and twin-based linkage studies has revealed the approximate location of numerous traits and disease conditions [7, 2]. Regions of potential linkage could be replicated only in some studies, however [1]. Ample patient and population based studies revealed associations of disease related phenotypes to patterns of SNPs marking regions of linkage disequilibrium that harbor the causal effects. In only a few studies, however, could the complex phenotype be assigned to the influence of common [11] or rare [6] functional gene variants. The genome-wide association scan (300,000–500,000 SNPs per subject) is currently the most complete assessment of genetic background that is still feasible in population studies [12, 13]. Such scans will identify common variants contributing to a trait but the effect of rare variants will be missed. In addition, the approach needs large sample sizes and various steps of replication to find true positive results because of the huge number of statistical tests that are being performed. Clearly, the hypothesis and explorative approaches in genetic studies each have their limitations.

Demonstrating replication of findings in genetic studies of complex traits has been difficult for a number of reasons [1] of which the (small) sample size of studies is the most frequent item of concern. In addition replication depends on the phenotype definition, patient registries, the size and heterogeneity of the genetic component. Improvement of this situation may be expected by the identification of functionally relevant SNPs in coding and regulatory

P. E. Slagboom (✉) · B. T. Heijmans · M. Beekman ·
I. Meulenbelt
Department of Molecular Epidemiology,
Leiden University Medical Centre, Research Building S5-P,
PO Box 9600, 2300 RC Leiden, The Netherlands
e-mail: p.slagboom@lumc.nl

Fig. 1 Two approaches to genome-wide scans for localization and identification of complex trait loci



sequences, by studying larger or joint cohorts, by meta-analyses, combined linkage association studies, improved molecular phenotyping [10], measuring environmental conditions and improving the statistical analysis tools. The analysis of gene–gene and gene–environment interactions in sufficiently large studies of extensively genotyped subjects is vital to an understanding of a complex trait [13].

Study example

In our studies we have integrated, for a number of research lines, the genetic and molecular profiling on well-phenotype hospital-based, twin and population-based cohorts. Genome-wide scans and candidate gene studies are performed simultaneously using various genotyping platforms each with their own range of optimal output. The phenotypes of subjects in these studies include clinical diagnosis, endpoints of disease, disease progression, intermediate phenotypes and advanced biomarker measurements. We will illustrate how candidate pathway and whole genome analyses can be integrated into the study designs we apply to human longevity and healthy ageing.

Although it is widely accepted that the major focus of genetic research into human health is by studying diseased subjects, extremely long-living subjects may be equally relevant. The delayed occurrence of disease and death from disease by the combined influences of genetic and lifestyle factors in long-living subjects may point at mechanisms beneficially influencing the ageing process and pathophysiology of major diseases. In the Leiden Longevity Study we have collected 430 nonagenarian siblings (93 years on average) enriched for familial influences on longevity [14] since their parents, siblings and children live longer than their birth cohort. In the sibling pairs we perform a genome-wide linkage scan to detect major longevity loci. To detect common variants

with smaller effects we perform a genome-wide association scan comparing nonagenarian siblings and controls: the partners of the children of the siblings (60 years on average). The first genome-wide scan for this phenotype revealed positive linkage on chromosome 4 and a candidate gene in the linkage area that associated to longevity [8]. In the Leiden Longevity Study, however, the linkage could not be replicated and a meta-analysis of the candidate longevity gene showed that it had been a false positive association [4].

Prominent phenotypes in human ageing research are disease-related parameters or the age at death of study subjects. Currently, the field is lacking biomarkers or intermediate phenotypes that can be used in genetic studies. Biomarkers would be parameters that provide inside into the biological age of subjects and their risk of dying rather than the chronological age. The design of the Leiden Longevity Study allows for the investigation of potential biomarkers of ageing and for a combined analysis of genetics and biomarkers. A whole spectrum of potential biomarkers is being investigated in the Leiden Longevity Study including cytokine profiles, NMR and mass spectrometry profiles of serum proteins and urinary metabolites, gene expression profiles, DNA methylation patterns and telomere shortening. In addition data on nutrition and lifestyle are collected. The first example of a potential biomarker of ageing is the low density lipoprotein (LDL) particle size in serum. We and others found that exceptionally old individuals have a remarkable lower number of small LDL particles [9, 3]. Small LDL particles are considered to be atherogenic [5]. Evidence from animal studies and trials support the importance of LDL size and also that these particles reflect metabolic rate. The beneficial lipoprotein particle profile observed in the nonagenarian siblings is already present among their middle aged children as compared to partners of the children, of which a propensity to become long-lived is expected. Studies

investigating whether and how these profiles protect from disease in middle age (the childrens' generation) may provide clues for the underlying biological mechanism.

In summary, genetic research of complex traits requires the inclusion of multiple study populations and designs, the complementary analysis strategies of linkage, association and DNA sequencing and the complementary knowledge of the biology of humans and animal models. In most of these studies the influence of environmental factors such as lifestyle and nutrition is hardly investigated. Such data would enable a more sophisticated analysis of the gene by environment interactions.

Acknowledgments Our scientific work is financially supported by a.o. IOP Genomics/Center, the Dutch Ministry of Health, Welfare and Sports, the Centre for Medical Systems Biology (CMSB), and the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NWO), Unilever.

References

1. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69:936–950
2. Arya R, Duggirala R, Almasy L, Rainwater DL, Mahaney MC, Cole S, Dyer TD, Williams K, Leach RJ, Hixson JE, MacCluer JW, O'Connell P, Stern MP, Blangero J (2002) Linkage of high-density lipoprotein-cholesterol concentrations to a locus on chromosome 9p in Mexican Americans. *Nat Genet* 30:102–105
3. Barzilai N, Atzmon G, Schechter C, Schaefer EJ, Cupples AL, Lipton R, Cheng S, Shuldiner AR (2003) Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA* 290:2030–2040
4. Beekman M, Blauw GJ, Houwing-Duistermaat JJ, Brandt BW, Westendorp RG, Slagboom PE (2006) Chromosome 4q25, microsomal transfer protein gene, and human longevity: novel data and a meta-analysis of association studies. *J Gerontol A Biol Sci Med Sci* 61:355–362
5. Berneis K, Rizzo M (2004) LDL size: does it matter? *Swiss Med Wkly* 134:720–724
6. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305:869–872
7. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM et al (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130–136
8. Geesaman BJ, Benson E, Brewster SJ, Kunkel LM, Blanche H, Thomas G, Perls TT, Daly MJ, Puca AA (2003) Haplotype-based identification of a microsomal transfer protein marker associated with the human lifespan. *Proc Natl Acad Sci USA* 100:14115–14120
9. Heijmans BT, Beekman M, Houwing-Duistermaat JJ, Cobain MR, Powell J, Blauw GJ, van der Ouderaa F, Westendorp RG, Slagboom PE (2006) Lipoprotein particle profiles mark familial and sporadic human longevity. *PLoS Med* 3:e495
10. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, Mueller M, Hummel O, Monti J, Zidek V, Musilova A, Kren V, Causton H, Game L, Born G, Schmidt S, Muller A, Cook SA, Kurtz TW, Whittaker J, Pravenec M, Aitman TJ (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet* 37:243–253
11. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugier L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 379:821–823
12. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PV, Frazer KA, Cox DR, Ballinger DG (2005) High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet* 77:685–693
13. Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37:413–417
14. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, Westendorp RG (2006) Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 14:79–84