THE JOURNAL OF
MATERNAL-FETAL
& NEONATAL
MEDICINE

http://informahealthcare.com/jmf
ISSN: 1476-7058 (print), 1476-4954 (electronic)

J Matern Fetal Neonatal Med, 2013; 26(S2): 9–12
© 2013 Informa UK Ltd. DOI: 10.3109/14767058.2013.830405

informa
healthcare

REVIEW PAPER

Twin, association and current ‘‘omics’’ studies

Dorret I. Boomsma

Department of Biological Psychology, VU University, Amsterdam, The Netherlands

Abstract

This paper describes the estimation of heritability, the genetic analysis of comorbid traits based on twin designs and on indices based on measured genetic relatedness. Next, approaches to identify genes and to examine the modification of heritability are described. The paper concludes with a discussion on the continued value of twin studies.

Keywords

Gene hunting, genetic epidemiology, omics, twin studies

History

Published online 19 September 2013

Introduction

Individual differences in growth, development, health and behavioral traits can in part be explained by genetic differences between individuals. When a trait is highly genetic, this indicates that a large proportion of its variance is explained by genetic factors, i.e. the effects of one or more genes that each influences the expression of the trait. Non-genetic factors can range from intrauterine and perinatal environment, to the influence of neighborhood, life events, friends and many other unidentified environmental factors, including random influences and ‘developmental noise’ [1]. This paper discusses approaches used to model the contribution of genes to variance in one or more traits, and to identify the regions of the genome that are involved. The focus is on methods that have evolved to analyze data from twins, but throughout the generalization to other data will be included. The last part discusses factors that influence the expression of genes, how gene expression is regulated and how genes interact with each other and with environmental exposures.

Genetic information is encoded in the DNA code, which contains the units of genetic information called *genes*. There is no strict consensus on what defines a gene; commonly used definitions include: ‘‘a unit of inheritance’’ or ‘‘a packet of genetic information that encodes a protein or RNA’’. The estimated number of genes in the human genome is also dynamic. Estimates have gone down from 100,000 genes to 20,000–50,000. In humans, the DNA molecules are organized in 22 pairs of *autosomes* and one pair of *sex chromosomes*. Two corresponding chromosomes are called *homologous* chromosomes, one copy is inherited from the mother, the other from the father. In addition, a small amount of DNA is contained in the maternally inherited *mitochondria.* The entire genetic sequence is called the *genome*, and a location in the genome that contains a gene or a genetic marker is referred to as a *locus*. *Quantitative trait loci (QTLs)* harbour genes influencing quantitative traits, i.e. a trait that varies on a quantitative scale, or the liability to a complex disorder, when the assessment of the trait is dichotomous (e.g. affected – unaffected).

Most of the human DNA sequence is identical in all individuals, but at some loci different versions of the sequence, called *alleles,* occur. The two alleles constitute a *genotype*. Individuals who carry the same allele at both homologous chromosomes are *homozygous*. Individuals with two different alleles are *heterozygous*. The term *haplotype* refers to a combination of alleles at multiple loci that an individual receives from one parent (usually a combination of alleles transmitted close together on the same chromosome). Finally, measurable characteristics of an individual are called *phenotypes*.

Alleles can affect a phenotype in various ways. When alleles act independently and their effects add up, they are additive. When the effect of one allele depends on the effect of another, there is genetic non-additivity. Interaction between two alleles at the same locus is referred to as genetic dominance, at different loci, it is referred to as epistasis. Interactions can also occur between genes (G) and environment (E): when the effect of a genotype depends on the environment, or when the impact of the environmental intervention depends on the genotype of a person this is described by GE interaction. When genes and environment are not independent, this is referred to GE correlation.

Address for correspondence: Dorret I. Boomsma, Department of Biological Psychology, VU University, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Tel: +31 20 598 8787. Fax: +31 20 598 88 32. E-mail: di.boomsma@vu.nl

Estimating heritability from twin- and other designs

The fact that many traits are familial is compatible with heritability, but familial resemblance can also result from the influence of a shared environment. On the other hand, there are some highly heritable traits, autism being one well-known example, that show little evidence for familial clustering, except in monozygotic twins. Genetic and shared environmental influences can be distinguished by studying adoptions. Similarities of adopted children and their biological relatives reflect genetic influences, whereas similarities of adopted children and their adoptive relatives reflect the effects of the shared environment. However, adoptions are relatively rare, and adoptive children and parents cannot be assumed to be representative of the general population. For this reason, many studies analyze data from twin families, in which the resemblance between monozygotic (MZ) and dizygotic (DZ) twins is compared to estimate the contribution of genes to the variance in a trait. MZ twins arise after a single ovulation, after the fertilized egg splits, and thus share 100% of their genes. DZ twins arise after a double ovulation and share on average 50% [Note that the 50% refers to the part of the genome in which variation or segregation occurs]. In contrast, MZ *and* DZ twins share their pre- and part of their postnatal environment. This implies that differences between MZ twins are due to non-shared environmental influences, whereas the extent to which MZ twins are more similar than DZ twins reflects the influence of genetic factors. The percentage of variance in a trait that is explained by additive genetic factors (A) equals the *narrow-sense heritability* (h^2) of a trait, which can be estimated by taking twice the difference between the MZ and DZ twin correlations: $h^2 = 2(rMZ - rDZ)$. When $rMZ > 2rDZ$, there is evidence for a contribution of non-additive genetic influences, also referred to as genetic dominance (D) and the variance explained by A and D together is referred to as the *broad-sense heritability* (H^2).

Often the contribution of genetic factors to a trait, or to the association (comorbidity) between multiple traits is estimated in *structural equation models*. Parameter values can be estimated using optimization approaches such as maximum likelihood and the goodness-of-fit of different models can be compared by the likelihood between models. Continuous variables are assumed to be normally distributed, which is indeed expected for traits that are affected by many genes and environmental effects. Clearly, non-continuous phenotypes (e.g. presence or absence of a disorder) are not normally distributed. However, they may reflect a categorization of an underlying normally distributed trait. In this situation, a liability model [2] is often used, which assumes that the categories of a variable reflect an imprecise measurement of an underlying liability. One or more thresholds divide this liability into discrete classes (e.g. 'affected versus unaffected'). The resemblance of relatives is expressed as a tetrachoric correlation, which stands for the correlation on the liability scale.

Recently, two new methods for the estimation of (narrow-sense) heritability were proposed that make use of measured genome-wide typed SNPs (single nucleotide polymorphisms) in large groups of unrelated subjects, rather than employing the theoretical values of genetic resemblance in relatives. The two approaches differ substantially, with one approach resembling the variance decomposition methods as used in twin studies [3], and the other [4] based on density estimation (DE) methods. One method [3] requires raw genotype data and uses these to obtain a measure of genetic similarity between all possible pairs of (unrelated) individuals in the study. In a second step, this genetic relatedness matrix (GRM) is used to predict the phenotype similarity between individuals (just as the different similarity of MZ and DZ twin pairs predicts their different phenotype resemblances). The DE method [4] can be applied after a genome-wide association study (GWAS, see below) has been done. Here, the distribution of z-statistics of the association measure between SNPs and the phenotype in a GWAS is compared to the theoretical Null distribution of z-statistics representing no effects. Explained variance will differ from zero if more z-statistics from the GWAS have larger values than expected under the Null. Heritability based on twin data compared to those based on GRM and DE methods for major depression, smoking and continuous measures of fasting glucose and height found that that a substantial proportion of the twin-based heritability estimates is recovered by the GRM and DE methods [5].

An important extension of the models described above is the analysis of multiple traits. By analyzing bivariate or multivariate data, the genetic and environmental overlap in correlated traits can be estimated, and the etiology of the association between traits explored, by testing whether the same genes affect correlated traits, or whether similar environmental factors are responsible for the correlation. Information regarding the correlation between traits comes from the *cross-twin cross-trait correlation* (the correlation between trait 1 in twin 1 and trait 2 in twin 2): if this cross-twin cross-trait correlation is higher in MZ than in DZ twins, this indicates that there is a genetic correlation between them. In a series of bivariate analyses the relationship between asthma, rhinitis and eczema in children was investigated [6], to test the hypothesis that the comorbidity of these disorders is due to shared genes. Genetic correlations ranged from 0.47 (asthma–rhinitis) to 0.62 (rhinitis–eczema), demonstrating that common genes play a role in more than one disease. It is noteworthy that the genetic correlations all were smaller than 1; thus, there were also genetic influences unique to each disease.

Gene finding

Since it is not-yet-feasible to characterize the entire human DNA sequence in large numbers of individuals, gene-finding studies rely on markers. Markers are genetic variants (*polymorphisms*) with known locations. When individuals are genotyped for linkage or association studies, their DNA is characterized at a large number of marker loci, either in a specific region (in candidate gene studies) or throughout the genome (in genome-wide studies). Several types of markers are used in gene-finding studies. *Single Nucleotide Polymorphisms* (*SNPs*) are single base pairs with two variants (e.g. some individuals have an A, others have a C). Theoretically (if single base pair mutations have occurred multiple times at the same locus) there can be four variants (A, C, T and G), but usually SNPs with two variants are selected for gene-finding studies. *Microsatellites* are sequence length polymorphisms that consist of a varying number of repeats of a short (1–4 base pairs) sequence of DNA, e.g. ''CACACACACACA''. A third type of polymorphism is the *copy number variant* (CNV). CNVs are DNA fragments ranging from kilobases (Kb) to even megabases (Mb) in size. The presence of discordance in CNVs has been demonstrated within MZ twin pairs who were selected for discordance for Attention problems and ADHD [7]. This implies that the assumption that MZ twins are genetically identical is not always fully correct, but also that discordance in MZ twins for highly genetic traits such as childhood ADHD can in sometimes be explained by their discordance for CNVs. We currently are at the beginning of an exciting new line research studying genetic dissimilarities within MZ twins pairs, as an alternative and effective gene-finding strategy [8].

Genetic association studies test whether a particular allele, genotype, or CNV is more prevalent in individuals with a certain phenotype: for instance, do individuals with allele C have more ADHD than those with allele A? Such association studies can be performed in unrelated individuals or in family-based samples. It is also possible to test for association with continuous phenotypes: in this case mean trait values are compared between individuals with different genotypes.

The EArly Genetics & Lifecourse Epidemiology (EAGLE) and the Early Growth Genetics (EGG) consortia have initiated a number of GWA projects in infants and children for birth weight, childhood obesity, growth, atopic dermatitis and other traits. A GWAS of 11 025 atopic dermatitis cases and 40 398 controls identified two new risk loci for atopic dermatitis near genes that have roles in epidermal proliferation and differentiation [9], supporting the importance of abnormalities in skin barrier function in the pathobiology of atopic dermatitis. In addition, a significant signal was seen from within the cytokine cluster at 5q31.1 (because of the large number of tests carried out in GWA studies, a p value of 0.5×10^{-8} is required to declare a finding genome-wide significant). Replicated signals were found for the epidermal differentiation complex, representing the FLG (filaggrin) locus, and a chromosome 11q13.5 variant. These results are consistent with the hypothesis that atopic dermatitis is caused by both epidermal barrier abnormalities and immunological features.

One concern that is sometimes raised in GWA studies is the inclusion of twins, as they more often are born prematurely and have lower birth weights. A comparison of twins to their siblings and to population standards for final height and body mass index (BMI) found that at 18 years, twins were as tall as their siblings, although they were leaner [10]. Compared to the general population, twins attained the same height and BMI. With respect to heritability, estimates based on data from twins and on singleton children and their parents from the same Dutch population, showed highly similar results. For length at birth heritability was 26% in singletons and 27% in twins. At 36 months, the estimates for height were 63% and 72%, respectively. Heritability estimates for birth weight were 26% in singletons and 29% in twins [11].

The outcome of large-scale GWA studies can be used to construct polygenic scores for each individual with genome-wide marker data in independent samples. Polygenic scores reflect the weighted sum of multiple SNP alleles associated with the trait (at a liberal significance threshold) and can be tested for the same phenotype as analyzed in the discovery GWA, or for traits hypothesized to be genetically related to it. In adults, this approach showed a genetic association between personality and major depression or bipolar disorder [12] and in children it was recently established that polygenic scores for ADHD are especially enriched in children with comorbid conduct disorder [13].

Modification of heritability

Heritability may differ for males and females, as a function of age or environmental exposures. Note that differences in heritability may arise for multiple reasons: the genetic variance can differ between the sexes, or in different age groups, or the genetic variance can be the same while the environmental variance is larger in men than in women or in older than in younger people. Since heritability is expressed as a ratio (genetic variance over total variance), these scenario's lead to different heritability estimates. Maybe general intelligence (IQ) is the best example of a trait whose heritability depends on age [14]. In children, the heritability of IQ is low; it increases during early puberty and reaches 70–80% in adults. Less well-known, and of opposite effect is the change in heritability for ADHD [15]. ADHD and attention problems are highly heritable in childhood (70% or more) while in adults heritability is only around 30–40%.

It is also possible, regardless of the size of heritability, that different genes affect the trait across the life-span or in males and females, although the empirical evidence for such sex-differences on a genome-wide basis is not strong [16].

When it is known which genes influence a phenotype, it is possible to test for interaction of the environment with a candidate gene (cGE interaction). In a critical review of cGE findings in psychiatry [17], many more novel cG × E studies turned out to be significant compared to replication attempts suggesting that cG × E hypotheses appear more robust than they actually are. Substantial progress has been made in studies of human complex traits, when the field moved from the study of candidate genes towards genome-wide methods and such methods should now be adapted for the study of GE interactions.

Conclusions

A person's phenotype depends on more than the genetic code. Genes exert their effects through their products and need to be expressed. The main steps in gene expression are *transcription* and *translation*. During transcription, DNA molecules serve as templates to construct RNA copies (RNA molecules resemble DNA but contain a different base, and are single-stranded). RNA codes for a sequence of amino acids, together forming a protein. Protein synthesis, based on the RNA code, is called translation. The expression of genes is affected by factors, such as epigenetic modifications [18] and regulation by other genes or *transcription factors* (proteins that bind to DNA, thereby controlling gene expression). Epigenetic modification of the DNA code can sometimes explain the discordance within MZ twin pairs for genetic disorders as described for the AXIN1 gene [19]. In an MZ pair discordant for a caudal duplication, this region was significantly more methylated in the affected than in the unaffected cotwin.

The study of discordant MZ twins emerges as a powerful method to identify genes, copy number variants, biomarkers and metabolites that are associated with disease. Van Dongen et al. [20] review multiple applications of the classical twin design, and the discordant MZ twin study and conclude that ‘‘classical twin methods combined with novel technologies represent a powerful approach towards identifying and understanding the molecular pathways that underlie complex traits’’.

Declaration of interest

This work was supported by the European Research Council (ERC Advanced Grant 230374).

References

1. Molenaar PC, Boomsma DI, Dolan CV. A third source of developmental differences. Behav Genet 1993;23:519–24.
2. Falconer DS, MacKay TFC. Introduction to quantitative genetics. 4th ed. Harlow, Essex, UK: Longmans Green; 1996.
3. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010;42:565–9.
4. So H-C, Li M, Sham PC. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet Epidemiol 2011;35:447–56.
5. Lubke GH, Hottenga JJ, Walters R, et al. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. Biol Psychiatry 2012;72:707–9.
6. Van Beijsterveldt CE, Boomsma DI. Genetics of parentally reported asthma, eczema and rhinitis in 5-yr-old twins. Eur Respir 2007;29: 516–21.
7. Ehli EA, Abdellaoui A, Hu Y, et al. De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur J Hum Genet 2012;20:1037–43.
8. Zwijnenburg PJ, Meijers-Heijboer H, Boomsma DI. Identical but not the same: the value of discordant monozygotic twins in genetic research. Am J Med Genet B Neuropsychiatr Genet 2010;153B: 1134–49.
9. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet 2012;44:187–92.
10. Estourgie-van Burk GF, Bartels M, Boomsma DI, Delemarre-van de Waal HA. Body size of twins compared with siblings and the general population: from birth to late adolescence. J Pediatr 2010; 156:586–91.
11. Mook-Kanamori DO, van Beijsterveldt CE, Steegers EA, et al. Heritability estimates of body size in fetal life and early childhood. PLoS One 2012;7:e39901.
12. Middeldorp CM, de Moor MH, McGrath LM, et al. The genetic association between personality and major depression or bipolar disorder. A polygenic score analysis using genome-wide association data. Transl Psychiatry 2011;1:e50.
13. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for adhd in children with comorbid aggression. Am J Psychiatry 2013;170:909–16.
14. Haworth CM, Wright MJ, Luciano M, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry 2010;15:1112–20.
15. Kan KJ, Dolan CV, Nivard MG, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry 2013;52:12–25.
16. Vink JM, Bartels M, van Beijsterveldt TC, et al. Sex differences in genetic architecture of complex phenotypes? PLoS One 2012;7: e47371.
17. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry 2011;168:1041–9.
18. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 2013;14:585–94.
19. Oates NA, van Vliet J, Duffy DL, et al. Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. Am J Hum Genet 2006;79:155–62.
20. van Dongen J, Slagboom PE, Draisma HH, et al. The continuing value of twin studies in the omics era. Nat Rev Genet 2012;13: 640–53.