

SHORT REPORT

Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained

Peter M Visscher^{*1}, Toby Andrew^{2,3} and Dale R Nyholt¹

¹Genetic Epidemiology, Queensland Institute of Medical Research, Herston, Brisbane, Australia; ²Department of Epidemiology & Public Health, Imperial College, St Mary's Campus, Norfolk Place, London, UK; ³Twin Research and Genetic Epidemiology Unit, St Thomas' Hospital, London, UK

For complex disease genetics research in human populations, remarkable progress has been made in recent times with the publication of a number of genome-wide association scans (GWAS) and subsequent statistical replications. These studies have identified new genes and pathways implicated in disease, many of which were not known before. Given these early successes, more GWAS are being conducted and planned, both for disease and quantitative phenotypes. Many researchers and clinicians have DNA samples available on collections of families, including both cases and controls. Twin registries around the world have facilitated the collection of large numbers of families, with DNA and multiple quantitative phenotypes collected on twin pairs and their relatives. In the design of a new GWAS with a fixed budget for the number of chips, the question arises whether to include or exclude related individuals. It is commonly believed to be preferable to use unrelated individuals in the first stage of a GWAS because relatives are 'over-matched' for genotypes. In this study, we quantify that for GWAS of a quantitative phenotype, relative to a sample of unrelated individuals surprisingly little power is lost when using relatives. The advantages of using relatives are manifold, including the ability to perform more quality control, the choice to perform within-family tests of association that are robust to population stratification, and the ability to perform joint linkage and association analysis. Therefore, the advantages of using relatives in GWAS for quantitative traits may well outweigh the small disadvantage in terms of statistical power.

European Journal of Human Genetics advance online publication, 9 January 2008; doi:10.1038/sj.ejhg.5201990

Keywords: genome-wide association; genetic association; quantitative trait; non-centrality parameter; statistical power

Introduction

Recent publications of genome-wide association scans (GWAS) for a range of diseases^{1–4} and quantitative phenotypes^{4–6} have demonstrated the feasibility of this

'unbiased' approach to gene discovery. It is now clear from published GWAS that effect sizes are small, with relative genotype risks typically <1.5. For quantitative traits, the individual effect sizes are consistent with <1% of the phenotypic variance being explained by a single polymorphism.^{4,5,7} For such traits, it might appear inefficient to include related individuals in the first stage of a GWAS because relatives are 'over-matched' for genotypes. It is known that for simple tests of association with disease, where the assumption of common causal variants is true,

*Correspondence: Dr PM Visscher, Genetic Epidemiology, Queensland Institute of Medical Research, 300 Herston Road, Herston, Brisbane 4029, Australia. Tel: +61 7 3362 0166; Fax: +61 7 3362 0101;

E-mail: peter.visscher@qimr.edu.au

Received 13 September 2007; revised 27 November 2007; accepted 28 November 2007

the use of relatives at the expense of unrelated individuals can cause a reduction in power. For example, sib-controls are over-matched to index cases, leading to a loss of power compared with unrelated case-control studies.^{8,9} Although most GWAS to date have used unrelated cases and controls, a number of studies have used related individuals.^{6,10}

Numerous association study designs for binary phenotypes were considered by Risch and colleagues^{8,9,11} and recently by others.¹² The conclusion from these studies was that selecting multiple cases from multiplex families increased power, in particular for a rare allele with a large effect on disease susceptibility, but that selecting controls that are related decreased power. The reason for the former is that the susceptibility allele is enriched in the multiplex families when there is a strong phenotype-genotype relationship. The reason for the latter is that the controls are over-matched relative to the cases.

However, common variants with small effect size that are targeted in a GWAS imply that for statistical power, both the advantages and disadvantages of having relatives diminish. Here we show, for a GWAS of a quantitative trait, that surprisingly little power is lost when genotyping related individuals. Since genotyping of relatives has many advantages (QC, linkage analysis, parent of origin effects), these results argue for including relatives in a GWAS where possible.

Methods

For a quantitative trait, we assume that the QTL heritability (q^2) is small (so that $1 - q^2 \approx 1$ and $\ln(1 + q^2) \approx q^2$) and assume an additive model. Let ρ be the phenotypic correlation of the relatives and r the coefficient of relationship (=twice the kinship coefficient). We consider the non-centrality-parameter (NCP, e.g.¹³) of a test for association, using either two unrelated individuals or a pair of related individuals with coefficient of relationship r . For n unrelated individuals, the NCP is $NCP_U = nq^2/(1 - q^2) \approx nq^2$, so $2q^2$ per pair of unrelated individuals. The NCP for any pair of relatives can be derived using regression theory. Following Visscher and Duffy,¹⁴ the NCP per family (size two) is, $NCP_{relatives} \approx 2q^2(1 - \rho r)/(1 - \rho^2)$. For sibships ($r = 1/2$), this result is identical to the approximate NCP for total association ($\lambda_B + \lambda_W$) from Sham *et al*¹³ The ratio of this NCP to that from having two unrelated individuals is,

$$NCP_{relatives}/NCP_U \approx (1 - \rho r)/(1 - \rho^2) \approx 1 - \rho(r - \rho) \quad (1)$$

This simple expression shows that the relative power of unrelated *versus* related pairs of individuals only depends on the phenotypic correlation of the relatives and the coefficient of relatedness. In practice, the phenotypic correlation will usually be smaller than the coefficient of relatedness. For example, for sibling pairs or dizygotic twin pairs ($r = 1/2$), estimates of phenotypic correlations for

systolic blood pressure and body mass index were 0.23¹⁵ and 0.26,¹⁶ respectively. For some traits, in particular those where the resemblance between relatives has a strong environmental component, sibling phenotypic correlations can be $> 1/2$. For example, estimates of sibling phenotypic correlations for leukocyte telomere length and forced expiratory volume were 0.67¹⁷ and 0.64,¹⁸ respectively. When the resemblance between relatives is solely due to additive genetic effects then the phenotypic correlation is always smaller than the coefficient of relatedness. In which case, the ratio in Equation [1] is less than one and power is lost by using relatives. For sibling pairs, the ratio of NCP is $(1 - (1/2)\rho)/(1 - \rho^2)$. For the special case that all family resemblance is due to additive genetic effects ($\rho = rh^2$), the ratio of NCP for pairs of relatives is $(1 - r^2h^2)/(1 - r^2h^4) \approx 1 - r^2h^2(1 - h^2)$. For sibships of arbitrary size, the results from Sham *et al*¹³ can be used to quantify the ratio of NCP. The ratio of the approximate NCP for s siblings *versus* s unrelated individuals was derived from,¹³

$$NCP_{sibs(s)}/NCP_{unrelated} = [1 - (1/2)\rho(s - 3)]/[(1 - \rho)(1 + \rho(s - 1))] \quad (2)$$

Results

We used the above results to quantify the loss in power for sibships (Figure 1) and, assuming an additive genetic model of family resemblance, for pairs of related individuals with a range of relationships (Figure 2). Clearly, the loss in efficiency is very small for small sibships and for pairs of relatives with a coefficient of relationship less than $1/2$. For sibling pairs, the largest loss (ratio of NCP of 0.93, or loss of 7%) is for a phenotypic correlation of 0.3 (Figure 1). For larger sibships, genotyping all siblings leads

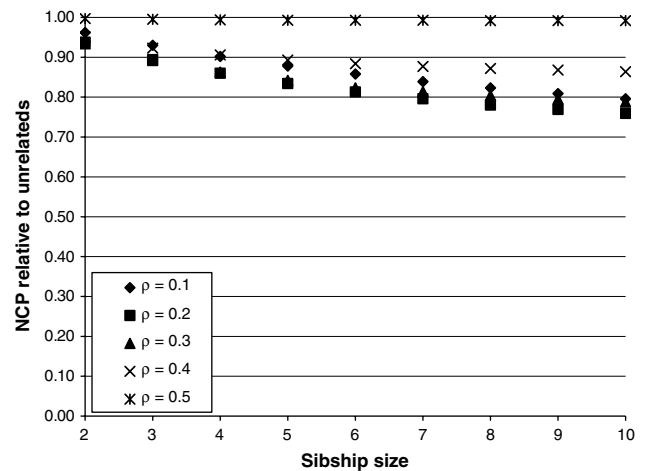


Figure 1 Relative power of GWAS for sibships *versus* unrelated individuals, for the same cost of genotyping. ρ is the phenotypic correlation between siblings.

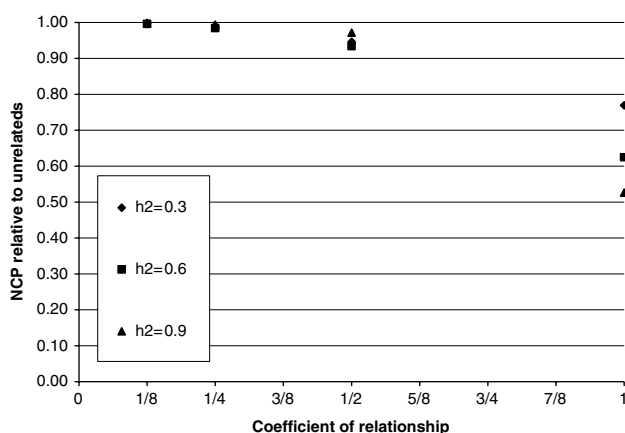


Figure 2 Relative power of GWAS for pairs of related individuals versus unrelated individuals, for the same cost of genotyping, assuming that all family resemblance is due to additive genetic effects. The coefficient of relationship (x axis) is twice the kinship coefficient. Value of 1/8, 1/4, 1/2 and 1 correspond to, for example, first cousins, half sibs, full sibs and monozygotic twin pairs.

to a theoretical loss of power of 5–20% for a realistic range of parameters (Figure 1). The maximum loss in power approaches 50% when genotyping both pairs of monozygotic twins ($r=1$) and when the heritability is large (Figure 2), but clearly both monozygotic individuals would not be genotyped in practice for a GWAS. By differentiating Equation [1], for a given coefficient of relationship (r) the loss in power is maximum when the phenotypic correlation (ρ) is approximately $r/2$ (since $[1-\sqrt{(1-r^2)}]/r \approx r/2$), and the relative power at this value of ρ is $\sqrt{(1-r^2)}/\{1-[1-\sqrt{(1-r^2)}]/r\}^2 \approx 1-r^2/4$. For sib pairs ($r=1/2$), the minimum power of 0.933 is obtained when ρ is 0.268. The quantified loss in power (say, 5–10%) is small relative to the loss in power in most GWAS due to incomplete SNP coverage.^{1–4}

In extreme cases, when the phenotypic correlation is larger than the coefficient of relationship, using relatives can actually increase power. For example, for telomere length and forced expiratory volume (assuming phenotypic correlations of 0.67 and 0.64 for sibling pairs), the efficiency of having pairs of siblings relative to the same number of unrelated individuals is 1.152 and 1.207, respectively (from Equation (1)).

Discussion and conclusions

The benefits of having relatives in a GWAS are manifold. They include the ability to perform more quality control, for example Mendelian error checking and IBD sharing, the choice to perform within-family tests of association that are robust to population stratification, the ability to perform parent-of-origin analysis and the ability to perform joint linkage and association analysis. Furthermore,

where cohorts of related individuals already exist, including large consortia such as GenomeUTwin,¹⁹ substantial gains in power may be obtained by utilising relatives due to the resulting increase in sample size. Robust and powerful statistical methods exist, which can incorporate familial relationships in association analysis of both quantitative^{20,21} and binary^{21–23} phenotypes.

For large sibships and in general larger pedigrees, it would be inefficient to genotype all individuals for all markers (Figure 1). In those cases, only genotyping a subset of the individuals and imputing genotypes for ungenotyped individuals would be more cost effective.^{14,24} The NCP for association in an arbitrary complex pedigree can be investigated numerically by calculating the variance of the best linear predictor from a linear mixed model, assuming known variance components.²⁵

In conclusion, contrary to common belief, there is hardly any loss in power when using relatives to conduct an association study and the advantages of using relatives in GWAS may well outweigh the small disadvantage in terms of statistical power, by providing a more robust and flexible strategy for analysis.

Acknowledgements

We thank David Goldgar and Bill Hill for helpful discussions and the referees for useful suggestions. This work was supported by Australian NHMRC Grants 389892, 339462 and 442915 and Australian Research Council Grant DP0770096.

Electronic-Database Information

GenomeUTwin <http://www.genomeutwin.org/>

References

- 1 WTCCC: Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
- 2 Sladek R, Rocheleau G, Rung J *et al*: A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
- 3 Scott LJ, Mohlke KL, Bonnycastle LL *et al*: A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
- 4 Saxena R, Voight BF, Lyssenko V *et al*: Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316**: 1331–1336.
- 5 Weedon MN, Lettre G, Freathy RM *et al*: A common variant of HMG2 is associated with adult and childhood height in the general population. *Nat Genet* 2007; **39**: 1245–1250.
- 6 Scuteri A, Sanna S, Chen WM *et al*: Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 2007; **3**: e115.
- 7 Frayling TM, Timpson NJ, Weedon MN *et al*: A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007; **316**: 889–894.
- 8 Teng J, Risch N: The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. *Genome Res* 1999; **9**: 234–241.

- 9 Risch N, Teng J: The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998; **8**: 1273–1288.
- 10 Moffatt MF, Kabesch M, Liang L *et al*: Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 2007; **448**: 470–473.
- 11 Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 2000; **405**: 847–856.
- 12 Li M, Boehnke M, Abecasis GR: Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am J Hum Genet* 2006; **78**: 778–792.
- 13 Sham PC, Cherny SS, Purcell S, Hewitt JK: Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000; **66**: 1616–1630.
- 14 Visscher PM, Duffy DL: The value of relatives with phenotypes but missing genotypes in association studies for quantitative traits. *Genet Epidemiol* 2006; **30**: 30–36.
- 15 Hottenga JJ, Whitfield JB, de Geus EJ, Boomsma DI, Martin NG: Heritability and stability of resting blood pressure in Australian twins. *Twin Res Hum Genet* 2006; **9**: 205–209.
- 16 Schousboe K, Willemsen G, Kyvik KO *et al*: Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries. *Twin Res* 2003; **6**: 409–421.
- 17 Andrew T, Aviv A, Falchi M *et al*: Mapping genetic loci that determine leukocyte telomere length in a large sample of unselected female sibling pairs. *Am J Hum Genet* 2006; **78**: 480–486.
- 18 Ferreira MA, O’Gorman L, Le Souef P *et al*: Variance components analyses of multiple asthma traits in a large sample of Australian families ascertained through a twin proband. *Allergy* 2006; **61**: 245–253.
- 19 Peltonen L: GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res* 2003; **6**: 354–360.
- 20 Abecasis GR, Cardon LR, Cookson WO: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
- 21 Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM: PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; **74**: 367–369.
- 22 Thornton T, McPeck MS: Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet* 2007; **81**: 321–337.
- 23 Goring HH, Terwilliger JD: Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 2000; **66**: 1310–1327.
- 24 Chen WM, Abecasis GR: Family-based association tests for genome-wide association scans. *Am J Hum Genet* 2007; **81**: 913–926.
- 25 Lynch M, Walsh B: *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates, 1998.