Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings

Peter M. Visscher[*], Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, Nicholas G. Martin

Genetic Epidemiology Group, Queensland Institute of Medical Research, Brisbane, Australia

The study of continuously varying, quantitative traits is important in evolutionary biology, agriculture, and medicine. Variation in such traits is attributable to many, possibly interacting, genes whose expression may be sensitive to the environment, which makes their dissection into underlying causative factors difficult. An important population parameter for quantitative traits is heritability, the proportion of total variance that is due to genetic factors. Response to artificial and natural selection and the degree of resemblance between relatives are all a function of this parameter. Following the classic paper by R. A. Fisher in 1918, the estimation of additive and dominance genetic variance and heritability in populations is based upon the *expected* proportion of genes shared between different types of relatives, and explicit, often controversial and untestable models of genetic and non-genetic causes of family resemblance. With genome-wide coverage of genetic markers it is now possible to estimate such parameters solely within families using the *actual* degree of identity-by-descent sharing between relatives. Using genome scans on 4,401 quasi-independent sib pairs of which 3,375 pairs had phenotypes, we estimated the heritability of height from empirical genome-wide identity-by-descent sharing, which varied from 0.374 to 0.617 (mean 0.498, standard deviation 0.036). The variance in identity-by-descent sharing per chromosome and per genome was consistent with theory. The maximum likelihood estimate of the heritability for height was 0.80 with no evidence for non-genetic causes of sib resemblance, consistent with results from independent twin and family studies but using an entirely separate source of information. Our application shows that it is feasible to estimate genetic variance solely from within-family segregation and provides an independent validation of previously untestable assumptions. Given sufficient data, our new paradigm will allow the estimation of genetic variation for disease susceptibility and quantitative traits that is free from confounding with non-genetic factors and will allow partitioning of genetic variation into additive and non-additive components.

Introduction

The theoretical basis for the resemblance between relatives due to genetic factors was developed by R.A. Fisher in a now famous and classic paper that reconciled Mendelian and biometrical genetics [1]. Following that theoretical basis, quantitative genetic parameters are estimated from the resemblance between different types of relatives by equating the observed phenotypic covariance to the degree of genetic relationship, which is estimated from pedigree data. The degree of relationship is usually expressed as the coefficient of kinship [2] or the additive coefficient of relationship [2,3]. In a non-inbred population, the coefficient of relationship is the expected proportion of alleles identical-by-descent (IBD) between relatives and determines the additive genetic covariance between a pair of relatives. Maximum likelihood (ML) methods and software have been developed to estimate genetic (co)variances in simple [4] and large complex pedigrees [5–7], for univariate and multivariate models. What all these methods have in common is that they estimate genetic parameters from observed variation between and within families, assuming an underlying model for causative components of variance [3]. For example, in twin studies it is commonly assumed that the variance between families is due to common environmental and additive genetic effects, and

that the variance within families reflects individual environmental effects (for monozygotic [MZ] pairs) or both individual environmental and additive genetic effects (for dizygotic [DZ] pairs).

In human populations, the interplay of genetic, environmental, and cultural factors that cause family resemblance is complex; and crucially, the ultimate separation of nature and nurture effects can generally not be tested empirically

* To whom correspondence should be addressed. E-mail: peter.visscher@qimr.edu.au

Synopsis

Quantitative geneticists attempt to understand variation between individuals within a population for traits such as height in humans and the number of bristles in fruit flies. This has been traditionally done by partitioning the variation in underlying sources due to genetic and environmental factors, using the observed amount of variation between and within families. A problem with this approach is that one can never be sure that the estimates are correct, because nature and nurture can be confounded without one knowing it. The authors got around this problem by comparing the similarity between relatives as a function of the exact proportion of genes that they have in common, looking only within families. Using this approach, the authors estimated the amount of total variation for height in humans that is due to genetic factors from 3,375 sibling pairs. For each pair, the authors estimated the proportion of genes that they share from DNA markers. It was found that about 80% of the total variation can be explained by genetic factors, close to results that are obtained from classical studies. This study provides the first validation of an estimate of genetic variation by using a source of information that is free from nature–nurture assumptions.

through controlled experiments. If the true (unknown) effects causing between-family variance deviate from the assumed model of family resemblance, then the resulting estimates of genetic parameters, and their estimated standard errors (SE), will be biased. This bias could be severe if strong assumptions are necessary to estimate genetic parameters. In the classical twin design, only three underlying parameters are estimated, and strong assumptions regarding the causes of familial resemblance are necessary. For example, the assumption that twin resemblance due to common environmental effects is the same for MZ and DZ pairs is often made. Although some of these assumptions can and have been tested empirically [8,9], the use of twin data to estimate heritability, in particular for traits such as cognitive function, has been controversial [10].

Until now, it has been impossible to exclude a possible confounding between genetic and non-genetic causes of family

resemblance. We propose an alternative approach to estimate genetic variance that is based upon the observed proportion of the genome that is shared by relatives and does not make any assumptions about the variation between families.

The actual genome-wide relationship, defined as the proportion of the genome that two relatives share IBD, varies around its expectation because of Mendelian segregation [11–14], except for MZ twins and parent-offspring pairs. We use the term "actual" throughout, but other possibilities are "realized relationships" or "the proportion of the genome-shared IBD." It is possible to estimate this relationship with the use of genetic markers. If these estimates are accurate, then it is, in principle, feasible to estimate genetic parameters within families, obviating the need for contentious assumptions about the sources of between-family variation.

In this study, we estimated heritability for height in humans without making any assumptions regarding the causes of resemblance between relatives. We present the relevant theory and estimate the heritability of height from collections of 3,375 full-sib pairs, using genome-wide estimates of actual additive genetic relationships. Bias and accuracy of our estimation approach was explored analytically and by computer simulation. Ours is the first example of an estimate of heritability in humans for which a possible confounding between nature and nurture can be excluded.

Results

Simulated Data

We first assessed bias and accuracy of the estimates of variance components from our method using simulation studies and analytical predictions (see Materials and Methods). Table 1 shows the empirical mean and SE of the ML estimate of the heritability from actual relationships between sibling pairs and statistical power, and their theoretical predictions, for a range of population parameters. As predicted by theory (see Materials and Methods), the SE of the estimates are large, unless the number of pairs is large

Table 1. Predicted and Empirical Power and Standard Errors of Heritability Estimation from Genome-Wide IBD Sharing between Sib Pairs

n	h^{2a}	f^{2b}	Simulations			Theory		
			Mean $(h^2)^c$	SE(h^2)	Power	Mean (h^2)	SE(h^2)	Power[d]
2,500	0.4	0.0	0.27	0.17	0.13	0.21	0.28	0.22
		0.2	0.41	0.28	0.28	0.40	0.39	0.27
	0.6	0.0	0.43	0.22	0.36	0.43	0.26	0.40
		0.2	0.60	0.29	0.58	0.60	0.34	0.56
	0.8	0.0	0.65	0.21	0.66	0.64	0.23	0.66
		0.2	0.70	0.16	0.92	0.80	0.27	0.90
10,000	0.4	0.0	0.31	0.13	0.52	0.31	0.14	0.52
		0.2	0.39	0.19	0.64	0.40	0.20	0.65
	0.6	0.0	0.52	0.12	0.87	0.51	0.13	0.87
		0.2	0.60	0.17	0.97	0.60	0.17	0.97
	0.8	0.0	0.73	0.11	1.00	0.72	0.11	0.99
		0.2	0.75	0.08	1.00	0.80	0.14	1.00

[a]Proportion of phenotypic variance due to additive genetic effects.
[b]Proportion of phenotypic variance due to non-genetic family effects.
[c]The mean and SE of the estimated heritability were calculated using equations given in the Materials and Methods.
[d]Power was calculated at 0.05 significance, using Purcell et al. [43].
DOI: 10.1371/journal.pgen.0020041.t001

(10,000), the heritability is large (>0.6), or there is no residual family effect. For 2,500 sib pairs, the SE of the heritability is approximately 0.2 when the true value is 0.8. For 10,000 sib pairs, the range of SE is from 0.08–0.19. The theoretical predictions are accurate, in particular for the special case when the proportion of variance due to non-genetic family effects (f^2) is zero. When the proportion of variance due to non-genetic family effects is zero, the estimate of the heritability is biased downwards, in particular when the sample size is small (Table 1). This is because we constrain variance components to be non-negative in our ML estimation procedure. An analytical prediction of the bias is given in Materials and Methods. When the heritability is large (0.8), its estimate is biased downwards, even when the proportion of variance due to non-genetic family effects was larger than zero. Again this is the result of ML estimation, because the sum of the proportion of variance due to genetic and non-genetic factors cannot be larger than unity.

Data Application

There were a total of 4,401 quasi-independent sibling pairs with estimates of genome-wide IBD sharing statistics. The average proportion of the genome-shared IBD between the sib pairs (the coefficient of additive genetic variance) was 0.498 (SE 0.0005, standard deviation [SD] 0.036), with a range of 0.374–0.617. The distribution of the genome-wide additive coefficients is shown in Figure 1. The mean and range of the proportion of the genome for which a sibling pair shared two alleles IBD (the coefficient of dominance variance, also termed IBD2) was 0.248 (SE 0.0006, SD 0.040) and 0.116–0.401, respectively. Hence, both mean sharing statistics were slightly lower than the expected values of 0.50 and 0.25,

respectively. When comparing the mean sharing statistics to their SE, there was evidence for a small but significant departure from expectation ($p = 0.002$ and 0.0002, for genome-wide additive and dominance coefficients, respectively, assuming a normal distribution of the test statistic). However, the SE is under-estimated because not all pair-wise sib comparisons are independent, so that the departure from expectation is less significant than it appears from the reported p-values. The SD of the mean (additive) IBD and mean IBD2 (dominance) sharing proportions were 0.036 and 0.040, respectively. One quality control measure of our IBD calculations is to test for independence of chromosome-specific additive and dominance relationships. For the combined dataset, 8/231 and 2/231 Spearman rank correlations of the mean IBD sharing between chromosomes were significant at the 0.05 and 0.01 level, respectively, when 12 and 2 were expected under the assumption of independent segregation. For IBD2 sharing the corresponding numbers were 9/231 and 1/231. The observed numbers are not significantly different from expectation under the null hypothesis of independent segregation of chromosomes (the SD of the number of significant correlations at the 0.05 and 0.01 level under the null hypothesis is 3.3 and 1.5, respectively).

Figure 2 shows the empirical variance of genome-wide mean IBD and IBD2 sharing, relative to the expected value from theoretical considerations (see Materials and Methods). There is a remarkably good agreement between theory and data, with a correlation between the theoretical and empirical SDs across chromosomes of 0.98 for both mean IBD sharing and mean IBD2 sharing. The correlation between mean IBD and mean IBD2 sharing for 4,401 pairs was 0.91, close to the theoretical value of 0.89 (Figure 3). This large correlation implies a strong sampling correlation between the estimates of additive and dominance variance.

ML estimators of heritability are shown in Table 2 for the two datasets separately and for the combined dataset. For each dataset, two models were fitted: a full model (FAE), containing a non-genetic family effect (F), a genome-wide additive effect (A), a residual error effect (E); and a reduced model containing F and E effects only (FE). In all analyses, the estimate of the residual family component was zero, and the estimate of heritability 0.8. For the combined dataset ($n = 3,375$ pairs), the 95% confidence interval (CI) was from 0.46 to 0.85 (a SE of approximately 0.1) with strong statistical support ($p = 0.0003$) for a variance associated with genome-wide IBD. The SE of the estimated proportion of variance due to additive genetic variance (h^2) is large relative to the estimate. However, because the sampling correlation of the estimates of the non-genetic and genetic variance is large and negative, the estimate of the total proportion of variance explained by genetic and non-genetic effects (i.e., the predicted MZ correlation) is more accurate. For the combined dataset, the ML estimate of this proportion is 0.80, with a 95% CI of 0.62–0.85. Hence, we have estimated the equivalent of an MZ correlation without having such pairs in our data.

The estimates from the FE model reflect the sibling correlation of 0.40 and 0.39 for the adolescent and adult datasets. Estimates of the proportion of variance due to additive genetic effects from the AE model (not shown in Table 2) were very close to twice the estimates of the
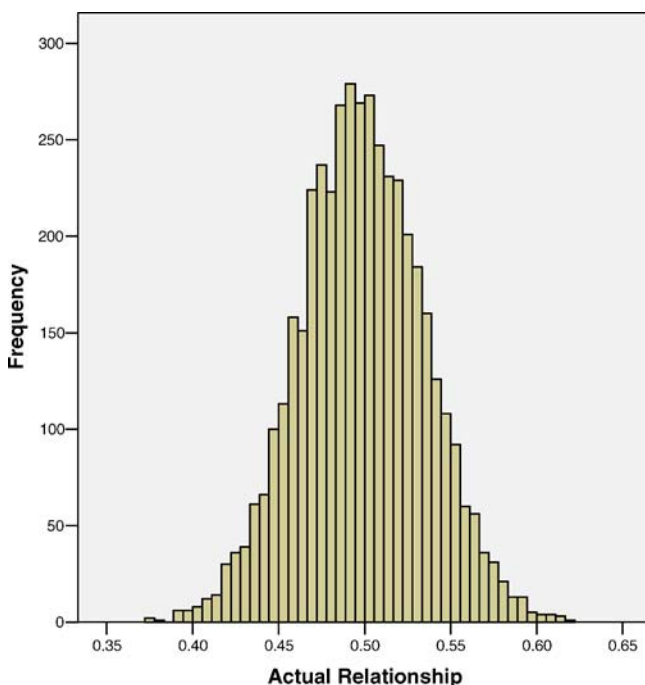

Figure 1. Empirical Distribution of Actual Additive Genetic Relationships of 4,401 Quasi-Independent Pairs of Full Sibs

Histogram of the genome-wide additive genetic relationships of full-sib pairs estimated from genetic markers.

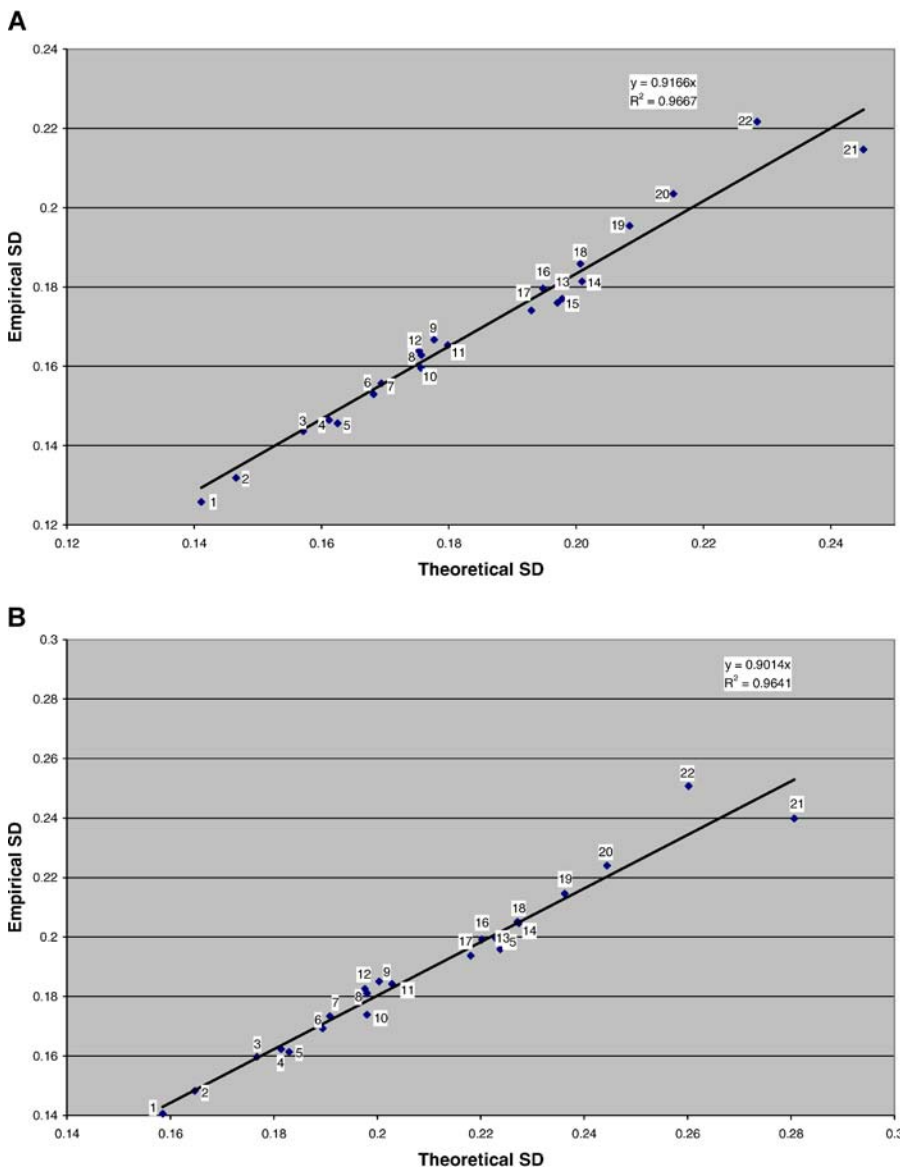DOI: 10.1371/journal.pgen.0020041.g001

Figure 2. Comparison of the Empirical and Theoretical Standard Deviations in Genome-Wide Sharing in 4,401 Quasi-Independent Pairs of Full Sibs
IBD sharing (A) and genome-wide IBD2 sharing (B) from marker data on 22 autosomes. The x- and y- axes are the theoretical and empirical SD, respectively. Numbers around the regression line indicate chromosomes.
DOI: 10.1371/journal.pgen.0020041.g002

proportion of variance due to the family effect in the FE model.

When genome-wide dominance was fitted in addition to F and A, the log-likelihood did not increase significantly for the combined dataset (unpublished data). However, there is unlikely to be sufficient power to distinguish these components with our sample size, consistent with our observed correlation coefficient of 0.89 between the additive genetic and dominance coefficients (Figure 3).

Discussion

We have shown that it is feasible to estimate genetic parameters solely from segregation within families, without making any assumptions regarding an underlying model for between-family effects. In fact, our only assumption in the

analysis is that the additive genetic covariance between relatives is proportional to the actual proportion of the genome that is shared IBD. The resulting estimates of the heritability for height (0.80, 95% CI, 0.46–0.85) and residual family effects (0.00, 95% CI, 0.00–0.17) are very close to estimates from twin studies [15], where the information comes from the difference in correlation between MZ and DZ twin pairs. Essentially, we have estimated the same parameters from DZ and full-sib pairs only.

Previously, methods have been proposed to estimate kinship and genetic parameters from marker data when pedigree data are not available, for example, in natural populations [16–18]. Relationship estimation and reconstruction in these methods are based upon identity-by-state sharing of marker alleles. These methods have the same principle as our approach, i.e., first estimating kinship from
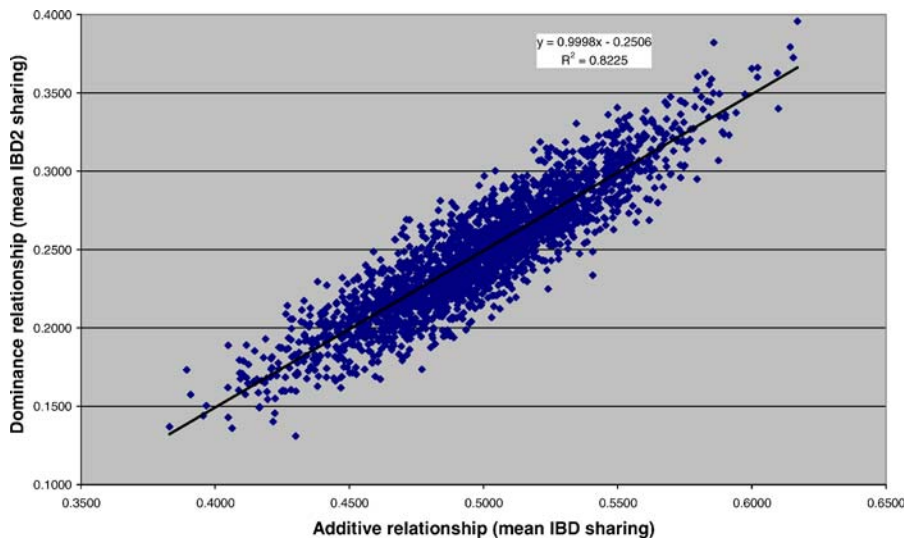
Figure 3. Correlation between Genome-Wide Mean IBD and Mean IBD2 Sharing

The x- and y- axes are the genome-wide additive and dominance relationships, respectively. Each point represents the genome-wide additive and dominance relationship for a sibling pair, estimated from genetic markers ($n = 4{,}401$ pairs).
DOI: 10.1371/journal.pgen.0020041.g003

marker data and subsequently estimating genetic variance from the association between phenotype similarity and estimated kinship. However, there are some important differences between the methods. Firstly, our method is based upon IBD sharing, i.e., we know the pedigree and estimated actual relationships from marker data conditional on the pedigree. The resulting estimates of actual kinship are unbiased and have lower error variance, provided that the pedigree is correct. Secondly, we estimate genetic variance free from possible confounding with environmental factors. In natural populations, even if the kinship were to be estimated without error, there can still be a confounding between genetic and environmental similarities and this could lead to bias.

We do not suggest that all estimation of genetic (co)variance from classical designs that utilize between-family comparisons should be abandoned. On the contrary, such designs, for example, those employing twin families, are in principle powerful enough to separate genetic and non-genetic causes of family resemblance if the statistical models are correct or at least a good approximation of the true

underlying causes of variation. With sufficient data, our approach allows the testing of hitherto untestable underlying assumptions in other models and, for large samples, allows the estimation of non-additive genetic variation for disease susceptibility and quantitative traits. Therefore, the two methods should be seen as complementary.

There is a continuum in the estimation of genetic parameters from genome-wide IBD sharing to quantitative trait loci (QTL) mapping. In QTL mapping, variation in IBD sharing is maximal but many estimations/tests are performed. For sib pairs, the variance of IBD sharing at a single location is $^1/_8$ [14,19–21], whereas it is only 0.039^2 genome-wide. Hence, relative to the mean there is about 82 times more variation in IBD sharing between sib pairs at a particular locus than in the genome-wide average [22]. The disadvantage of QTL mapping is that a genome-wide search is performed at many correlated locations, whereas the estimation of genetic variance from genome-wide IBD sharing is a single estimate. An intermediate between the two is to estimate the proportion of additive genetic variance associated with a chromosome [23–25]. The variance in proportion of a

Table 2. ML Estimates of Heritability of Height from Genome-Wide IBD Sharing between Sib Pairs

Data	Model	Estimates (95% CI)		LRT[a]	p-Value[b]
		f^2	h^2		
Adolescents ($n = 931$)	FAE	0.00 (0.00–0.43)	0.80 (0.00–0.90)		
	FE	0.40 (0.34–0.45)		1.850	0.0869
Adults ($n = 2{,}444$)	FAE	0.00 (0.00–0.18)	0.80 (0.43–0.86)		
	FE	0.39 (0.36–0.43)		9.817	0.0009
Combined ($n = 3{,}375$)	FAE	0.00 (0.00–0.17)	0.80 (0.46–0.85)		
	FE	0.39 (0.36–0.42)		11.553	0.0003

[a]Likelihood ratio test statistic for the null hypothesis that $h^2 = 0$, calculated from the difference in log-likelihood between models FAE and FE.
[b]p-Value calculated assuming that the LRT is distributed as zero with a probability of ½ and $\chi_{(1)}^2$ with a probability of ½.
LRT, likelihood ratio test.
DOI: 10.1371/journal.pgen.0020041.t002

Table 3. Theoretical Standard Deviation of IBD Sharing between Sibs per Chromosome

Chromosome	Length (Mb)[a]	Length (Morgan)[a]	SD(π)	Equivalent Number of Segments[b]
1	245.13	2.865	0.1411	6.28
2	243.16	2.633	0.1466	5.82
3	198.96	2.251	0.1571	5.06
4	191.51	2.122	0.1612	4.81
5	180.32	2.082	0.1625	4.73
6	169.96	1.922	0.1682	4.42
7	158.06	1.89	0.1694	4.36
8	145.7	1.733	0.1757	4.05
9	135.81	1.687	0.1777	3.96
10	134.6	1.735	0.1756	4.05
11	134.08	1.638	0.1798	3.87
12	131.43	1.742	0.1753	4.07
13	94.45	1.289	0.1978	3.19
14	85.13	1.238	0.2009	3.10
15	79.69	1.302	0.1971	3.22
16	89.51	1.342	0.1948	3.29
17	81.18	1.375	0.1929	3.36
18	75.33	1.242	0.2007	3.10
19	62.78	1.122	0.2084	2.88
20	61.38	1.025	0.2153	2.70
21	33.22	0.685	0.2451	2.08
22	33.11	0.861	0.2284	2.40
Total	2,764.5	35.781		84.79

[a]From [41].
[b]The number of independent loci for which the mean variance in IBD sharing would be equivalent to the observed variance for the whole chromosome (= 1/[8var(π)]).
DOI: 10.1371/journal.pgen.0020041.t003

chromosome-shared IBD is intermediate between the sharing proportion at a single location and genome-wide, as shown in Table 3. We note that the emphasis in this study is on the estimation of genetic parameters rather than its detection. Hence, in contrast to QTL mapping where hypothesis testing and p-values are important, we have concentrated on the sampling variance of the estimated parameters, because for most traits it is usually known that there is genetic variance, and the scientific question is what proportion of observed variation is genetic.

Although our estimates of the variation in mean IBD and IBD2 sharing per chromosome are very similar to the theoretical values (Figure 2) and consistent with recently reported genome-wide sharing statistics from a sample of 498 sib pairs [22], a few caveats are required. Firstly, the theoretical value may be too low for the true variance in IBD sharing on a chromosome because in reality there may be more crossovers than modeled [13,14]. Secondly, the empirical variance of IBD sharing is likely to be an underestimate because the marker information was not perfect. If we assume that our genome-wide average multipoint marker information content was approximately 80%, then we would expect to find a regression slope of the empirical on theoretical SD in IBD sharing of $\sqrt{0.80} = 0.89$, close to the observed value of 0.92 (Figure 2A). Nevertheless, the correlations of 0.98 between empirical and theoretical values are extremely high.

We detected a small genome-wide deviation of the observed IBD sharing statistics from expectation. Genome-wide transmission distortion, which results in excess allele sharing between relatives, has been reported previously [26].

Our results were driven by a deficit of the probability of sharing two alleles IBD, hence we do not replicate the findings of [26] with our large sample of 4,401 pairs.

Our simulation studies confirmed that a large number of pairs is needed for accurate estimation, and showed that the estimates of heritability were biased downwards when there was no underlying source of non-genetic family resemblance. This bias is the result of ML estimation because of the usual constraints that estimated variance components have to be non-negative and that the sum of the partitioned variance ratios is bounded by zero and one. The observed bias is not particular to our method because it applies to any variance partitioning approach by ML, in particular when sampling variances are large [27,28].

We have estimated a single additive genetic variance from genome-wide segregation of marker loci within families, after adjusting phenotypes for the fixed effects of sex and age at measurement. However, genetic variances for males and females and younger and older siblings may be different, and the genetic correlation across these groups may be smaller than unity. Although we have ignored these potential sources of heterogeneity of genetic variance in this study because of sample size considerations, models that include, for example, sex-limitation effects are, in principle, straightforward to implement.

We have ignored the contribution of the sex chromosomes to genome-wide IBD. In humans, the X chromosome accounts for 4% of genes and 5% of physical length [29]. If all chromosomes account for genetic variation in proportion to the number of genes or physical length, then our estimate of heritability will be biased downwards by about 4% to 5%.

Although our sample size of 3,375 was sufficient to estimate the heritability of height with reasonable accuracy, for phenotypes with smaller heritability (and to distinguish additive from dominance variance), larger sample sizes are necessary. Such large datasets are in the process of being generated, either from large national studies or by combining samples across countries. For example, the GenomEUtwin study will accrue over 10,000 sib pairs for linkage studies [30]. Therefore, in the near future we will be able to estimate unbiased genetic parameters for traits that have been controversial in the past due to the assumptions regarding the (non-genetic) resemblance between relatives. If a large population resource of relatives with measured phenotypes were to be available, then a selective genotyping strategy in which only concordant and discordant pairs are genotyped may be efficient in estimating quantitative genetic parameters accurately, for the same reason that such a design can be powerful in gene mapping studies [31,32].

Our application was on a single quantitative trait and using a simple pedigree structure. However, the method is entirely general and can be applied to disease phenotypes, multiple traits, and large arbitrary pedigrees. All that is required is genome-wide estimates of IBD sharing between relatives, observations on relevant phenotypes, large samples, and software to estimate components of (co)variance.

There are limitations of the applied method, the main one being that large sample sizes are required with dense marker coverage of genotyped individuals. This may be unachievable for most single labs now, but future large population-based studies that have a family component, or pooling of sample resources across studies, will have the desired effect of

increasing sample size. A second limitation is that sufficient markers need to be genotyped to obtain an accurate estimate of genome-wide sharing statistics. This is less of a problem because many samples that are suitable for our suggested analyses are genotyped for linkage studies, and marker density is likely to increase in the near future because of the availability of relatively cheap single nucleotide polymorphism genotyping. With the advent of high density single nucleotide polymorphism genotyping platforms, the error in estimation of genome-wide IBD sharing between relatives is likely to be small, and we have assumed, in the present study, that it is negligible. If the estimation of genome-wide IBD sharing is less than 100% accurate, then the variation in IBD sharing between pairs is less than the true variation, resulting in less powerful analysis but still unbiased estimates [33]. With less complete marker coverage, the estimate of the proportion of alleles shared IBD is unbiased $(E(\pi|\hat{\pi}) = \hat{\pi})$ but has larger prediction error variance. For a single location in the genome, we derived the prediction error variance as: $\mathrm{var}(\pi|\hat{\pi}) = \frac{1}{4}P_1(1-P_1) + P_2(1-P_2) - P_1P_2$, with P_i the probability of having i alleles IBD; note that this variance could be used as a weight in gene mapping studies. To a first order approximation, the sampling variance of the estimate of the heritability, relative to the situation of perfect marker information, is increased by the reciprocal of the average genome-wide information content [34]. A third limitation is that it is difficult to disentangle additive from non-additive effects. However, with sufficient data the large correlation between additive and dominance coefficients is not an issue, and one could even consider estimating additional non-additive effects, for example additive-by-additive or additive-by-dominance effects.

In conclusion, we have shown that it is feasible to estimate genetic variance entirely within families, by correlating phenotypes and genome-wide similarity. Our assumption-free method facilitates a complete separation of genetic and environmental causes of family resemblance and will allow the estimation and testing of non-additive sources of variation.

Materials and Methods

Variance of genome-wide IBD sharing. The variance of the proportion of chromosome segments that are IBD between relatives has been derived by a number of authors for pairs of full sibs [13,14,23,33,35], complex pedigrees [12,36,37], for inbred individuals [38], and for experimental backcross populations [39,40]. In the case of full sibs we give a derivation for both the additive and dominance component of covariance, and their correlation, following the approach of Hill [39].

Additive effects. For a given sib pair, the genome-wide mean IBD sharing (π) is the sum of the proportion shared from the paternal (p) and maternal (m) contribution,

$$\pi = \frac{1}{2}\pi^p + \frac{1}{2}\pi^m, \text{ and } \mathrm{var}(\pi) = \frac{1}{4}\mathrm{var}(\pi^p) + \frac{1}{4}\mathrm{var}(\pi^m). \quad (1)$$

Hence, to calculate the variance it is sufficient to consider the contribution from a single parent only. For parent k, the sharing of alleles by progeny depends on the proportion of alleles shared due to the parent's paternal or maternal gamete. Let δ_i be an indicator variable for locus i, which is one if both sibs have inherited the paternal allele or both sibs have inherited the maternal allele, and zero otherwise. Then,

$$E(\delta_i) = \frac{1}{2} \text{ and } \mathrm{var}(\delta_i) = \frac{1}{4}. \quad (2)$$

The covariance of the indicator variables at two loci (i and j) is:

$$\mathrm{cov}(\delta_i, \delta_j) = E(\delta_i\delta_j) - E(\delta_i)E(\delta_j) = \frac{1}{4}[1 + (1-2r)^2] - (\frac{1}{2})^2$$
$$= \frac{1}{4}(1-2r)^2. \quad (3)$$

Assuming the Haldane mapping function, the covariance can be written as:

$$\mathrm{cov}(\delta_i, \delta_j) = \frac{1}{4}\exp(-4\mathrm{d}_{ij}), \quad (4)$$

with d_{ij} the distance (in Morgan) between the loci. For n loci, the variance of chromosome-wide sharing between two sibs is:

$$\mathrm{var}(\pi^k) = (1/n^2)(\frac{1}{4})\Sigma\Sigma\exp(-4\mathrm{d}_{ij}) \quad (5)$$

(following [39]). If n becomes very large this equation can be expressed as an integral [12,39], $\mathrm{var}(\pi^k) = \frac{1}{4}\int_0^l\int_0^l e^{-4|x_1-x_2|}dx_1dx_2 = \frac{1}{8l^2}(l - r_{2l}/2)$ with l the length of the chromosome (in Morgan) and r_{2l} the recombination fraction for a segment of length $2l$. Hence, the total variance in IBD sharing between two siblings for chromosome i of length l is:

$$\mathrm{var}(\pi_i) = 2(\frac{1}{4})\mathrm{var}(\pi^k) = (16l^2)^{-1}(l - \frac{1}{2}r_{2l}). \quad (6)$$

Finally, genome-wide π is, $\pi_g = (1/L) \Sigma(l_i \pi_i)$, with $L = \Sigma(l_i)$, and:

$$\mathrm{var}(\pi_g) = (1/L^2)\Sigma l_i^2(16l_i)^{-1}(1 - \frac{1}{2}r_{2l_i}/l_i)$$
$$= [1/(16L^2)][L - \frac{1}{2}\Sigma r_{2l_i}]$$
$$\approx 1/(16L) - 1/(3L^2), \quad (7)$$

because there are 22 autosomes and $r_{2li} \approx \frac{1}{2}$. These results are the same as those of Guo [11], whose derivations were based upon Markov chains. They imply, that to a first order approximation, the variance in genome-wide IBD sharing is a function of the total genome length only [12,36,38,39]. For $L = 35$ Morgan, the SD of genome-wide IBD sharing is approximately 0.039. Table 3 shows a breakdown in the variance of IBD sharing per chromosome and the equivalent number of independent loci. It was constructed using the above equations, with physical and genetic lengths from [41], and using the sex-averaged recombination map. For comparison, the SD of the proportion of alleles shared at a given locus is 0.354.

Dominance. Dominance variance is a function of the probability that two siblings share both alleles IBD (= IBD2). In a non-inbred population, this probability is also called the coefficient of fraternity [2]. The prior probability that full sibs share two alleles IBD is $\frac{1}{4}$, and the mean and variance of an indicator variable that is one if both alleles are shared IBD and zero otherwise is $\frac{1}{4}$ and $\frac{3}{16}$, respectively. Note that the variance of IBD2 sharing at a single locus is 1.5 times the variance of mean IBD sharing. The probability that the sibs share two alleles IBD at a linked locus, given that they are IBD2, is $(1-r)^4 + 2[(1-r)r]^2 + r^4 = [(1-r)^2 + r^2]^2$. Hence the covariance of the indicator variable (δ) at loci i and j is:

$$\mathrm{cov}(\delta_i, \delta_j) = E(\delta_i\delta_j) - E(\delta_i)E(\delta_j) = \frac{1}{4}[(1-r)^2 + r^2]^2 - 1/16. \quad (8)$$

After some algebra it can be shown that the variance of the mean IBD2 sharing (π_{di}) on a chromosome of length l is:

$$\mathrm{var}(\pi_{di}) = [1/(16l^2)][(5/4)l - \frac{1}{2}r_{2l} - (1/16)r_{4l}]. \quad (9)$$

The genome-wide variance in mean IBD2 sharing is:

$$\mathrm{var}(\pi_d) = [1/(16L^2)][(5/4)L - \frac{1}{2}\Sigma r_{2l_i} - (1/16)\Sigma r_{4l_i}]$$
$$\approx 5/(64L) - 99/(256L^2) \approx 5/(64L) - 1/(3L^2). \quad (10)$$

Hence, the variance of the genome-wide IBD2 sharing is larger (by about 30% if L = 35) than the variance of the genome-wide mean IBD sharing. The correlation between mean genome-wide allele sharing and mean genome-wide IBD2 sharing is the ratio of the SD,

$$r(\pi_a, \pi_d) = \sigma(\pi_a)/\sigma(\pi_d) \approx [1/(16L)/\{5/(64L)\}]^{0.5} = 0.89. \quad (11)$$

The actual relationship between full sibs can be estimated with genetic markers. For fully informative markers and close relatives, only a few markers are needed per chromosome to capture the proportion of alleles shared IBD [23,24]. This is because the number of recombination events per chromosome is small.

Sampling variance of estimators of genetic variance. For n sib pairs, the simplest estimation procedure is to apply Haseman-Elston regression analysis [33] of the squared difference between the phenotypes (Y_{i1} and Y_{i2}) of the ith pair of siblings on the estimate of their genome-wide IBD proportion (π_i),

$$(Y_{i1} - Y_{i2})^2 = \alpha + \beta\pi_I. \quad (12)$$

The parameter β is proportional to the within-family additive genetic variance, adjusted for inbreeding in the parents,

$$\beta = -2\sigma_A^2[1 - \frac{1}{2}F_{mum} - \frac{1}{2}F_{dad}] \quad (13)$$

[2,3,33]. We will assume that parents are not inbred, so that the

regression slope equals minus twice the additive genetic variance. Then, an estimate of the narrow sense heritability is simply,

$$\hat{h}^2 = -\hat{\beta}/(2\hat{\sigma}_p^2) \tag{14}$$

with $\hat{\sigma}_p^2$ an estimate of the total phenotypic variance. If we ignore the sampling correlation between the estimate of the regression coefficient and the total phenotypic variance, then the sampling variance of the heritability is, using a Taylor series expansion [2]:

$$\text{var}(\hat{h}^2) \approx (\tfrac{1}{4})[\beta/\sigma_p^2]^2[\text{var}(\hat{\beta})/\beta^2 + \text{var}(\hat{\sigma}_p^2)/\sigma_p^4]$$
$$= h^4[\text{var}(\hat{\beta})/(4\sigma_p^4 h^4) + 1/n]. \tag{15}$$

The variance of the regression coefficient is approximately,

$$\text{var}(\hat{\beta}) \approx \left[8(1-t)^2/(n\text{var}(\pi) - 4h^4/n)\right]\sigma_p^4, \tag{16}$$

with t the sib intra-class correlation [20,21]. Hence, the sampling variance of the estimate of the narrow sense heritability is, approximately,

$$\text{var}(\hat{h}^2) \approx 2(1-t)^2/[n\text{var}(\pi)]. \tag{17}$$

This is fully analogous to the estimation of the proportion of variance explained by a single QTL, the only difference being the variance in genome-wide IBD sharing. The non-centrality-parameter (NCP) for a test of significance of genome-wide additive genetic variance is,

$$\text{NCP} = nh^4\text{var}(\pi)/[2(1-t)^2] \tag{18}$$

which reduces to the form given by Sham and Purcell [20] and Visscher and Hopper [21] for a single QTL when $\text{var}(\pi) = 1/8$. Following the derivations in Sham and Purcell [20] and Visscher and Hopper [21], the SE of the estimate of the heritability and NCP when using both the squared difference and squared sum of the sib pairs are, approximately,

$$\text{var}(\hat{h}^2) \approx (1-t^2)^2/[(1+t^2)\{n\text{var}(\pi)\}] \tag{19}$$

and

$$\text{NCP} = nh^4\text{var}(\pi)(1+t^2)/(1-t^2)^2. \tag{20}$$

Hence, power calculations for QTL mapping can be used to assess the sample size required to "detect" genome-wide additive genetic variance. For example, to detect a heritability of a given size is equivalent to detecting a QTL at a fully informative locus explaining q^2 of the phenotypic variance when $h^4 \, \text{var}(\pi) = (\tfrac{1}{8})q^4$, i.e., a QTL explaining about $0.11h^2$ of the phenotypic variance.

ML estimation uses more information than the difference between the sib pairs, and the resulting estimate of the heritability is more accurate. For a single QTL asymptotically (large sample size and a QTL that explains a small amount of variance), the sampling variance of the ML estimator is that of the least squares estimator, when both the squared differences and sums are used in the regression analysis [20,21]. For genome-wide estimation, the proportion of variance explained by π is small ($\sim 0.11h^2$), so it seems reasonable to use the predictions for the regression analysis. However, the predictions differ dramatically if there is no other source of family resemblance than sharing of genetic effects. The following approximate results were derived assuming the simple equation:

$$\hat{t}_0 = \hat{f}^2 + 0.5\hat{h}^2 \tag{21}$$

with $\hat{t}_0$ the estimate of the intra-class correlation under the null hypothesis of no genome-wide additive genetic variance, $\hat{f}^2$ the estimate of the proportion of the variance due to residual familial effects under the alternative hypothesis, and $\hat{h}^2$ the estimate of the heritability under the full model. Equation 21 is a good approximation because the intra-class correlation, which is estimated relatively precisely under the reduced model, is essentially partitioned into a genetic and non-genetic component in the full model. The sampling correlation between the estimates of f^2 and h^2 is approximately -1. If there are no constraints imposed on the estimates, then, using results from [42],

$$\text{var}(\hat{t}_0) \approx (1-t^2)^2/n \tag{22}$$

$$\text{var}(\hat{h}^2) \approx (1-t^2)^2/[n\,\text{var}(\pi)(1+t^2)] \tag{23}$$

(from [20,21]). By difference,

$$\text{var}(\hat{f}^2) \approx \text{var}(\hat{t}_0) + \tfrac{1}{4}\text{var}(\hat{h}^2) \approx \tfrac{1}{4}\text{var}(\hat{h}^2). \tag{24}$$

Hence, the SE of the estimate of the non-genetic familial resemblance is approximately half of the SE of the estimate of the genome-wide heritability. The above SE of the estimates can be used to calculate the probability that the ML estimate is zero, using standard normal distribution truncation theory [3] with truncation values of $-f^2/\sigma(\hat{f}^2)$ and $-h^2/\sigma(\hat{h}^2)$, respectively. This was validated using simulations (unpublished data).

Conditional on $f^2 = 0$. When the true residual familial component is zero, the ML estimate is zero with a probability of ½, and > 0 with a probability of ½ [27,28]. When the estimate of $f^2 = 0$ then the estimate of the heritability is approximately twice the intra-class correlation of the sibs. Hence, asymptotically,

$$\hat{h}^2 | \hat{f}^2 = 0 \sim N(h^2, 4(1-t^2)^2/n). \tag{25}$$

When the estimate of $f^2 > 0$, the mean estimate of the familial component is, approximately,

$$E(\hat{f}^2 | \hat{f}^2 > 0) = i\sigma(\hat{f}^2) \tag{26}$$

with i the mean value of a truncated standard normal distribution. For a truncation value of 0, as is the case here, $i = 0.798$ [3]. The variance of the truncated distribution is:

$$\text{var}(\hat{f}^2 | \hat{f}^2 > 0) = (1-i^2)\text{var}(\hat{f}^2). \tag{27}$$

Taking the whole of the distribution of the estimate of f^2 gives the mean and variances as:

$$E(\hat{f}^2 | f^2 = 0) = 0.5i\sigma(\hat{f}^2) = 0.4\sigma(\hat{f}^2), \tag{28}$$

and

$$\text{var}(\hat{f}^2 | f^2 = 0) = 0.5(1 - 0.5i^2)\text{var}(\hat{f}^2) = 0.34\text{var}(\hat{f}^2) \tag{29}$$

Similarly for the estimate of the heritability,

$$E(\hat{h}^2 | f^2 = 0) \approx h^2 - i\sigma(\hat{f}^2) \tag{30}$$

and

$$\text{var}(\hat{h}^2 | f^2 = 0) \approx 0.5(1 - 0.5i^2)\text{var}(\hat{h}^2) = 0.34\text{var}(\hat{h}^2). \tag{31}$$

Equations 23, 30, and 31 were used to predict the mean and SE of the estimate of the heritability and were found to be close to simulation results for large samples (Table 1). For small samples the distribution of the estimates of the two variance ratios could be approximated by a truncated bivariate distribution. This situation is more complex because the probability that either estimate is zero as well as the probability that the estimates are constrained at unity needs to be considered jointly. If there is no residual non-genetic family resemblance then the SE of the estimate of the heritability is nearly halved relative to the case where such effects are present. The case of no residual family resemblance is very unlikely for QTL mapping (where the effects of genes elsewhere in the genome and common environmental effects cause resemblance) but realistic for genome-wide analysis of highly heritable phenotypes. The reduction in SE is at the expense of a downward bias in the estimate of the heritability.

Models. The basic additive genetic model, fitted in both the simulation study and data application, is $Y_{ij} = \mu + F_i + A_{ij} + E_{ij}$, with μ the fixed effects of the mean and F, A, and E the random effects of non-genetic family, additive genetic, and residual factors, respectively. The covariance between the phenotypes of two siblings is modeled as $\text{cov}(Y_{i1}, Y_{i2}) = \text{var}(F_i) + \text{cov}(A_{i1}, A_{i2}) = \sigma_F^2 + \pi_{a(i)}\sigma_A^2$, and $\text{cov}(Y_{ij}, Y_{kl}) = 0$ if $i \neq j$. Extensions to non-additive models are straightforward, in principle. For example, the covariance for a model containing dominance (D) and additive-by-additive (AA) effects is: $\text{cov}(Y_{i1}, Y_{i2}) = \sigma_F^2 + \pi_{a(i)}\sigma_A^2 + \pi_{d(i)}\sigma_D^2 + \pi_{a(i)}\pi_{a(i)}\sigma_{AA}^2$.

Simulation. Simulations were performed to validate the predictions of the sampling variance of the heritability and statistical power. Genome-wide IBD sharing between pairs of sibs and their phenotypes were simulated from a simple model,

$$\pi_i \sim N(0.5, 0.04^2), Y_{ij} = \mu + F_i + A_{ij} + E_{ij}, \tag{32}$$

with μ, F, A, and E defined as before, with distributions

$$F_i \sim N(0, f^2), A_{i1} \sim N(0, h^2), A_{i2} | A_{i1} \sim N(\pi_{a(i)}A_{i1}, (1-\pi_{a(i)}^2)h^2) \text{ and}$$
$$E_{ij} \sim N(0, 1 - f^2 - h^2).$$

$$\tag{33}$$

Regression and ML analyses were performed (for details, see [21]). The number of pairs (n) in the simulation was either 2,500 or 10,000; heritability values were 0.4, 0.6, and 0.8; and the proportion of variance due to non-genetic family effects was either 0.0 or 0.2. For each set of population parameters, 1,000 replicates were run. Power was calculated using Web-based software for power of QTL analysis [43] at a type-I error rate of 0.05, which is appropriate because we performed a single hypothesis test.

Application to data. We estimated the mean and variance of genome-wide IBD sharing from 4,401 quasi-independent full-sib pairs, and applied the ML estimation method to 3,375 quasi-independent full-sib pairs with both marker data and phenotypic measurements on height. These data were collected from two cohorts of Australian twins and their siblings. Phenotypes for the adolescent cohort were collected in the context of continuing longitudinal studies examining risk factors for melanoma [44] and cognitive functioning [45]. For this cohort, height was measured during a clinical examination using a stadiometer at ages 12, 14, and 16; the most recent measurement being used in the current analyses. In the first instance phenotypes for the adult cohort (consisting of twins registered with the Australian Twin Registry born prior to 1971) were collected from self-report questionnaires. Through their subsequent participation in a variety of studies, 58% of the twins included here attended a clinical examination in which height was measured using a stadiometer [15,46]; self-reported height was analyzed if no clinical measurement existed. Correlation between clinically measured and self-reported height was 0.92 in individuals measured both ways [15]. Age at time of measurement was used as a covariate in both cohorts.

Genotypic information was available for a subset of the adolescent and adult participants. For the adolescent cohort, genotypic information was available for 1,201 individuals from 500 families, yielding 950 quasi-independent full-sib pairs. Genotypic information was available for up to 791 autosomal markers. The number of markers per participant in the current study ranged from 211 to 791, with a mean and SD of 588 and 194, respectively, giving an average marker spacing of 6 cM per genotyped individual. The genotyping, error checking, and cleaning of these data have been described in detail elsewhere [47]. For the adult cohort, genotypic information was available for 3,804 individuals from 1,512 families, yielding 3,451 quasi-independent full-sib pairs. Genotypic information was available for up to 1,717 autosomal markers. The number of markers per participant in the current study ranging from 201 to 1,717, with mean and SD of 628 and 264, respectively, and the average marker spacing was 5.6 cM per individual. Details of the genotyping, error checking, and cleaning strategies of these data are given elsewhere [48].

Phenotypes for height were missing on 481 individuals, eight in the adolescent cohort and 473 in the adult cohort. The number of sib pairs for which both individuals had a measured phenotype for the adolescent cohort, the adult cohort, and the combined cohort was 931, 2,444, and 3,375, respectively.

IBD probabilities at 1 cM intervals were calculated using Merlin [49], and the estimate of chromosome and genome-wide IBD sharing was enumerated by averaging the IBD probabilities over the length of a chromosome and the whole genome, respectively.

Each dataset was first adjusted for fixed effects, using a general linear model in which sex was fitted as a fixed factor and age at measurement as a linear covariate. Residuals from this analysis were standardized by the residual variance for each dataset because there was some evidence of heterogeneity of variance: the residual SD for the adolescent and adult dataset was 7.71 cm and 6.89 cm, respectively.

ML analysis was performed using Mx [4]. The full model, termed FAE, contained F and A and E. The covariance between the phenotypes of sibs one and two of pair i was modeled as $\text{cov}(Y_{i1}, Y_{i2}) = \sigma_F^2 + \pi_{a(i)}\sigma_A^2$, with $\pi_{a(i)}$ the estimate of the genome-wide actual additive relationship of the sibling pair. Reduced models FE and AE were subsequently fitted. A likelihood-ratio-test was performed to test the null hypothesis that A was zero, by comparing the MLs of models FAE and FE. A p-value was calculated assuming that the test statistic has an asymptotic distribution that is 0 with a probability of ½ and a one degree of freedom χ^2 with a probability of ½ [27,28]. CIs of the variance ratios were calculated by Mx and verified by a profile likelihood approach, in which one variance component at a time was changed from its ML value, while maximizing the likelihood for the remaining parameters, until a drop in twice the log-likelihood of 3.84 was reached. In addition to estimating the ML estimate of the variance components for F, A, and E, the ML estimate of (F + A) and its 95% CI were estimated. This was performed because the estimates of F and A have a large negative sampling correlation, so that the estimate of their sum is more precise than the estimate of the individual components.

References

1. Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans Roy Soc Edin 52: 399–433.
2. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland (Massachusetts): Sinauer Associates. 980 p.
3. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. London: Longman Press. 180 p.
4. Neale MC, Boker SM, Xie G, Maes HH (2003) Mx: Statistical modeling. Richmond (Virginia): Virginia Institute for Psychiatric and Behavioral Genetics.
5. Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. Biometrika 58: 545–555.
6. Lange K, Westlake J, Spence MA (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. Ann Hum Genet 39: 485–491.
7. Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440–1450.
8. Hettema JM, Neale MC, Kendler KS (1995) Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. Behav Genet 25: 327–335.
9. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993) A test of the equal-environment assumption in twin studies of psychiatric illness. Behav Genet 23: 21–27.
10. Devlin B, Daniels M, Roeder K (1997) The heritability of IQ. Nature 388: 468–471.
11. Guo SW (1996) Variation in genetic identity among relatives. Hum Hered 46: 61–70.
12. Stam P (1980) The distribution of the fraction of the genome identical-by-descent in finite random mating populations. Genet Res 35: 131–155.
13. Risch N, Lange K (1979) Application of a recombination model in calculating the variance of sib pair genetic identity. Ann Hum Genet 43: 177–186.
14. Suarez BK, Reich T, Fishman PM (1979) Variability in sib pair genetic identity. Hum Hered 29: 37–41.
15. Silventoinen K, Sammalisto S, Perola M, Boomsma DI, Cornes BK, et al.

(2003) Heritability of adult body height: A comparative study of twin cohorts in eight countries. Twin Res 6: 399–408.

16. Thomas SC (2005) The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. Philos Trans R Soc Lond B Biol Sci 360: 1457–1467.

17. Ritland K (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. Mol Ecol 9: 1195–1204.

18. Lynch M, Ritland K (1999) Estimation of pair-wise relatedness with molecular markers. Genetics 152: 1753–1766.

19. Rijsdijk FV, Hewitt JK, Sham PC (2001) Analytic power calculation for QTL linkage analysis of small pedigrees. Eur J Hum Genet 9: 335–340.

20. Sham PC, Purcell S (2001) Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. Am J Hum Genet 68: 1527–1532.

21. Visscher PM, Hopper JL (2001) Power of regression and maximum likelihood methods to map QTL from sib-pair and DZ twin data. Ann Hum Genet 65: 583–601.

22. Gagnon A, Beise J, Vaupel JW (2005) Genome-wide identity-by-descent sharing among CEPH siblings. Genet Epidemiol 29: 215–224.

23. Goldgar DE (1990) Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 47: 957–967.

24. Visscher PM (1996) Proportion of the variation in genetic composition in backcrossing programs explained by genetic markers. J Hered 87: 136–138.

25. Schork NJ (2001) Genome partitioning and whole-genome analysis. Adv Genet 42: 299–322.

26. Zollner S, Wen X, Hanchard NA, Herbert MA, Ober C, et al. (2004) Evidence for extensive transmission distortion in the human genome. Am J Hum Genet 74: 62–72.

27. Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82: 605–610.

28. Stram DO, Lee JW (1994) Variance components testing in the longitudinal mixed effects model. Biometrics 50: 1171–1177.

29. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. (2005) The DNA sequence of the human X chromosome. Nature 434: 325–337.

30. Peltonen L (2003) GenomEUtwin: A strategy to identify genetic influences on health and disease. Twin Res 6: 354–360.

31. Risch N, Zhang H (1995) Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science 268: 1584–1589.

32. Risch NJ, Zhang H (1996) Mapping quantitative trait loci with extreme discordant sib pairs: Sampling considerations. Am J Hum Genet 58: 836–843.

33. Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2: 3–19.

34. Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sib-ship data. Am J Hum Genet 66: 1616–1630.

35. Guo SW (1994) Computation of identity-by-descent proportions shared by two siblings. Am J Hum Genet 54: 1104–1109.

36. Guo SW (1995) Proportion of genome shared identical by descent by relatives: Concept, computation, and applications. Am J Hum Genet 56: 1468–1476.

37. Guo SW (1996) Gametogenesis processes and multilocus gene identity by descent. Am J Hum Genet 58: 408–419.

38. Franklin IR (1977) The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. Theor Popul Biol 11: 60–80.

39. Hill WG (1993) Variation in genetic composition in backcrossing programs. J Hered 84: 212–213.

40. Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30: 227–238.

41. Kong X, Murphy K, Raj T, He C, White PS, et al. (2004) A combined linkage-physical map of the human genome. Am J Hum Genet 75: 1143–1148.

42. Fisher RA (1921) On the "probable error" of a coefficient of correlation deduced from a small sample. Metron 1: 3–32.

43. Purcell S, Cherny SS, Sham PC (2003) Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19: 149–150.

44. Zhu G, Duffy DL, Eldridge A, Grace M, Mayne C, et al. (1999) A major quantitative-trait locus for mole density is linked to the familial melanoma gene CDKN2A: A maximum-likelihood combined linkage and association analysis in twins and their sibs. Am J Hum Genet 65: 483–492.

45. Wright M, De Geus E, Ando J, Luciano M, Posthuma D, et al. (2001) Genetics of cognition: Outline of a collaborative twin study. Twin Res 4: 48–56.

46. Healey SC, Kirk KM, Hyland VJ, Munns CF, Henders AK, et al. (2001) Height discordance in monozygotic females is not attributable to discordant inactivation of X-linked stature determining genes. Twin Res 4: 19–24.

47. Zhu G, Evans DM, Duffy DL, Montgomery GW, Medland SE, et al. (2004) A genome scan for eye color in 502 twin families: Most variation is due to a QTL on Chromosome 15q. Twin Res 7: 197–210.

48. Cornes BK, Medland SE, Morley KI, Ferreira MAR, Gordon S, et al. (2005) Sex-limited genome-wide linkage scan for body mass index in an unselected sample of 933 Australian Twin Families. Twin Res Hum Genet 8: 616–632.

49. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30: 97–101.