Like Mother, Like Daughter: Analysis of Parent-Child Phenotypic Correlations for Hundreds of Phenotypic Traits

Emma Pierson, Aaron Kleinman, Nicholas Eriksson and David Hinds

23andMe, Inc., Mountain View, CA, USA

Introduction

Children's phenotypes are strongly correlated with their parents' for many medical and behavioral phenotypes, because of a combination of genetic and environmental factors. Quantifying the contribution of each of these factors has remained an open area of research, and these calculations are typically performed by grouping parents together. Here, we compute parent-child phenotype correlations for more than 1,000 traits on the 23andMe cohort, and discover that maternal traits tend to be more correlated with daughters' phenotypes than paternal traits. Although many of the positively associated phenotypes are biological, we also found significant positive correlations for behavioral and personality traits. To quantify the genetic contribution to these associations, we computed the heritability of many of these phenotypes.

Overview

Cohort

- Participants were drawn from more than 600,000 genotyped customers of 23andMe, Inc. who consented to participate in research as part of the 23andMe Personal Genome Service®.
- Participants were genotyped on the Illumina HumanHap550+ and HumanOmniExpress+ platforms.
- Phenotypic information was gathered on a variety of behavioral, medicinal and personality traits through voluntary completion of web-based survey questions.

Methodology

- We phased participants using a modified version of BEAGLE [1] and computed pairwise IBD using GERMLINE [2].
- Parent-child relationships were detected by finding pairs of participants who shared more than 85% half IBD and less than 10% full IBD.
- Using this, we extracted a cohort (the "trio cohort") of 14,222 pairwise-disjoint father-mother-child trios of European ancestry.
- We stratified this cohort by the gender of the child and gender of the parent, and used regression to compute the effects of maternal and paternal phenotypes on sons and daughters (**Figure 1**).
- We triophased 8,116 children and computed parental-specific allelic associations at known imprinted genes (**Figure 2**).
- For 45 quantitative phenotypes of interest, we computed a Galton heritability by regressing the child's phenotype against the parental average (**Figure 3**).
- For 110 quantitative and binary phenotypes, we estimated heritability by running GCTA [3] on a cohort of more than 30,000 unrelated individuals of European descent (**Figure 4**).

Results

- Daughters' traits are more significantly affected by their mothers' traits than their fathers' traits; a similar result does not hold for sons.
- SNPs in imprinted genes appear to exhibit parent-of-origin effects.
- Heritability estimates from GCTA and from midparent regression correlate well, and midparent regression estimates are larger than the GCTA estimates. This can be used to measure environmental heritability.
- A large number of interesting social traits exhibit nonzero genetic heritability.

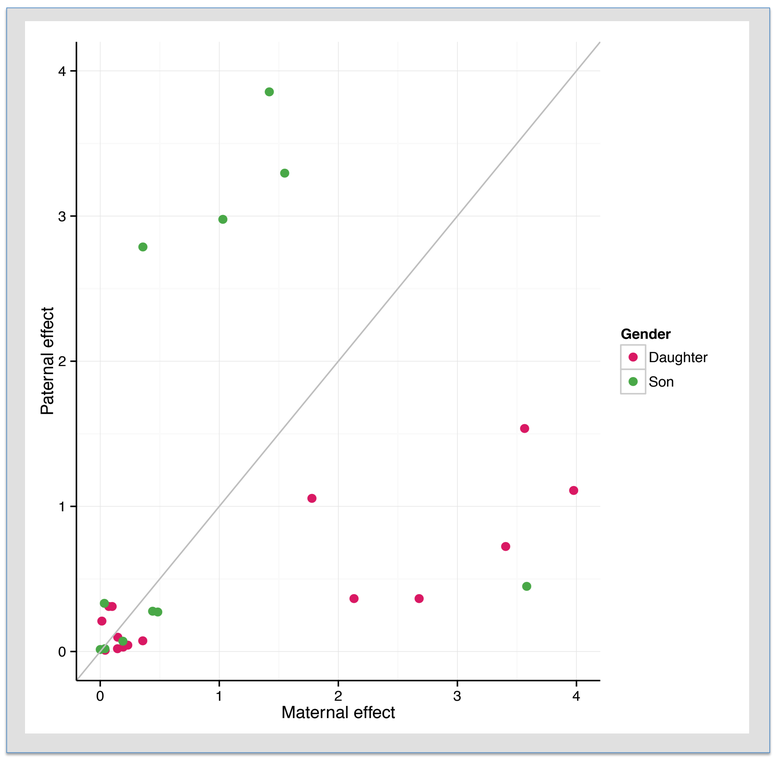

Figure 1. We stratified the trio cohort by the gender of the child and the gender of the parent. Then for each phenotype, we regressed the child's phenotype on the parents' phenotypes, controlling for the child's age and first five genetic principal components. Of the 1,066 phenotypes, 378 were statistically significant across one of the four subcohorts. 100% of statistically significant mother-daughter, mother-son and father-son effects were positive, and 99.5% of father-daughter effects were positive. We then computed an F-test statistic to determine when, for a given gender of child, the maternal and paternal effect sizes were different. For sons, the maternal effect size is greater in 11 of 20 phenotypes. For daughters, the maternal effect size is greater in 29 of 37 (p=7.5E-4). The gray line indicates the diagonal.

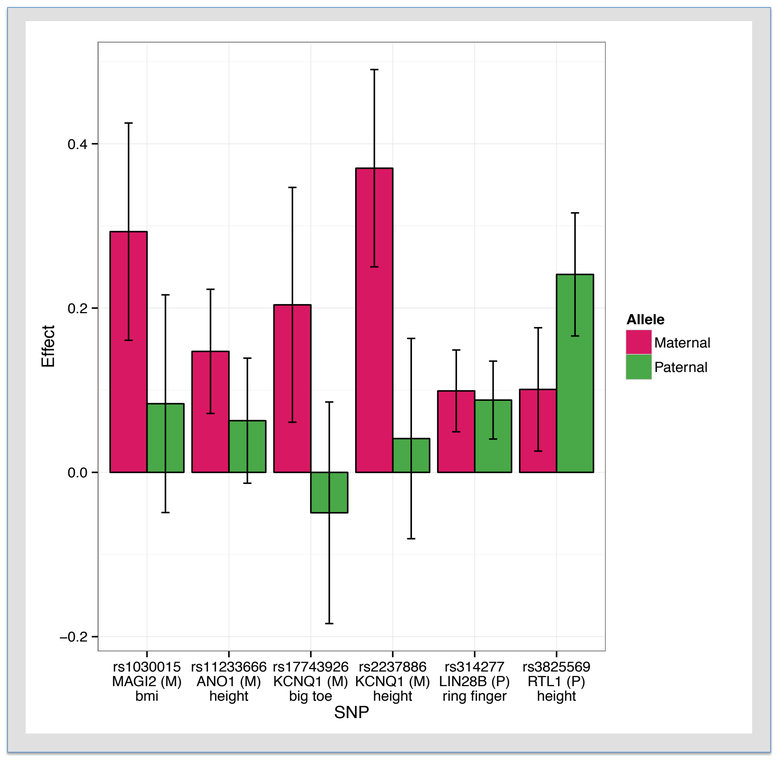

Figure 2. We examined GWAS results for 110 traits on a cohort of over 300,000 participants of European descent. For each phenotype, we considered the set of associations with p-value < 1E-6, and then controlled for LD by merging nearby associations and choosing the most significant association from each block. We then subset to those SNPs which appear in a list of 95 genes that are known to be imprinted [4]. This resulted in a list of 10 associations. Next, we selected a set of 8,116 European trios and used the parents to phase the children. As a crude proxy for the power of this cohort, we recomputed the known GWAS associations on the child cohort and discarded four with a p-value > 0.5. For each of the remaining six associations, we computed the parental allelic effect by regressing the child's phenotype against the child's age, sex, first five principle genetic components, and the maternal and paternal alleles, which were treated as separate covariates. The figure shows their effect sizes: The first four SNPs lie in known maternally imprinted genes (M), the last two in paternally imprinted genes (P). Error bars are 68% CI. Though the figure is suggestive, for each association, an F-test was unable to reject the null hypothesis that the parental effects differed.

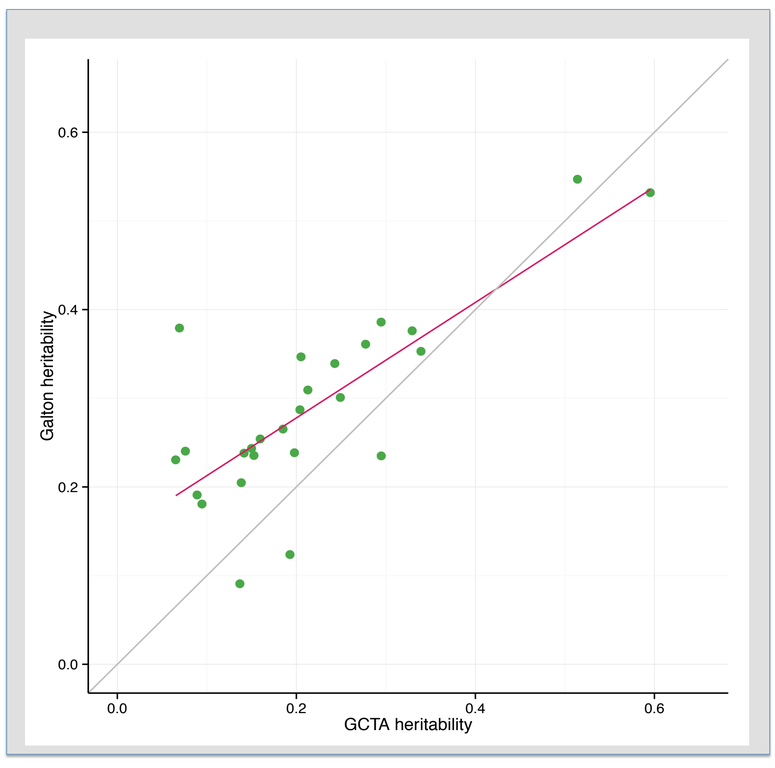

Figure 3. We computed estimates of heritability for 40 quantitative phenotypes in two ways. For each phenotype, we first estimated the effect size of age, sex and the first five genetic principal components by linear regression on a cohort of more than 300,000 participants of European descent. We then subtracted this off and, for the trio cohort, regressed the child's residual against the average of the parents' residuals to give a Galton heritability estimate. Next, we ran GCTA on a cohort of over 30,000 individuals of European descent. This figure shows the heritability estimates for phenotypes whose Galton estimate was significantly different from zero. Most of the phenotypes had larger Galton estimates than GCTA estimates, as expected because the Galton estimate also includes the effect of shared environment and additive-by-additive interactions. The pink line is a regression, the gray line is the diagonal.

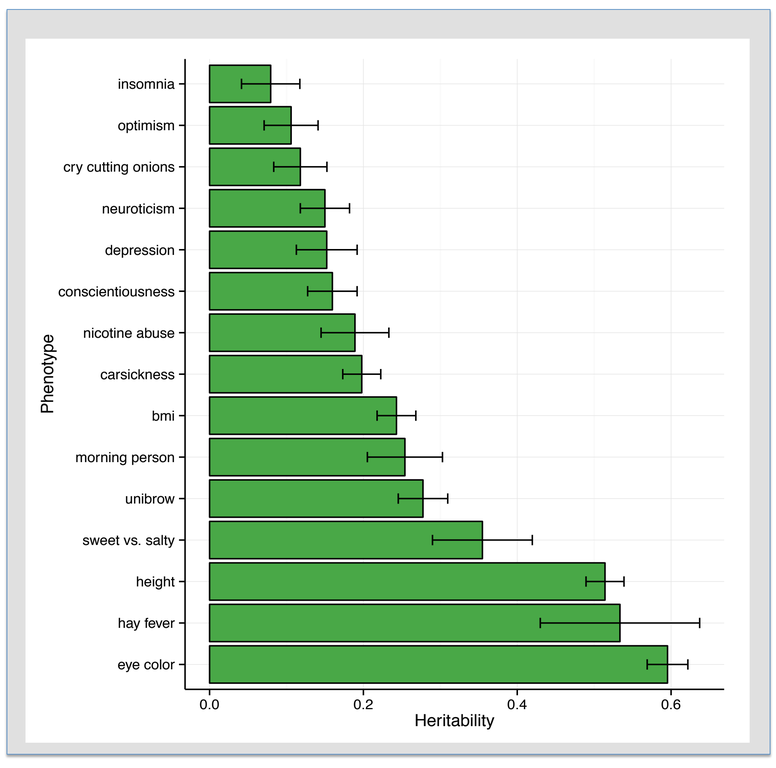

Figure 4. We ran GCTA on a cohort of more than 30,000 unrelated individuals of European ancestry to estimate the narrow-sense heritability of more than 100 phenotypes. To convert binary phenotypes to the liability threshold, prevalences were computed over the entire 23andMe European cohort. Error bars show 95% CIs.

Acknowledgments

We thank the customers of 23andMe who consented to participate in research and who answered surveys, as well as the employees of 23andMe, who together made this research possible.

References

[1] S R Browning and B L Browning (2007). *Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering.* Am J Hum Genet 81:1084-1097.

[2] Gusev A, Lowe JK, Stoffel M, Daly MJ, Altschuler D, Breslow JL, Friedman JM, Pe'er I (2008). *Whole population genomewide mapping of hidden relatedness.* Genome Research.

[3] Yang J, Lee SH, Goddard ME and Visscher PM. *GCTA: A Tool for Genome-wide Complex Trait Analysis*. Am J Hum Gent. 2011 Jan 88(1): 76-82.

[4] Randy Jirtle. www.geneimprint.com.