# DNA Revolution and the Social and Behavioral Sciences

MACIEJ TRZASKOWSKI and ROBERT PLOMIN

## Abstract

A century of genetic research on the social and behavioral sciences has addressed the "how much" question, showing that genetic differences are important for nearly all traits. However, during the past few decades, research has moved beyond this rudimentary "how much" question to ask "how" questions about developmental change and continuity, about the relationship between traits, and about the interplay between genes (nature) and environment (nurture). We suggest that some of the most important and transformative findings in the social and behavioral sciences have come from this research. Moreover, the most revolutionary changes in genetic research are on their way with the rapid advances in DNA technology and methodology, which promise to transform the social and behavioral sciences. It is crucial that social and behavioral scientists stay on top of the DNA revolution. The purpose of our essay is to provide an overview of genetic research in the social and behavioral sciences.

## NATURE AND NURTURE: TWIN AND ADOPTION STUDIES

In general, genetic research can be divided into quantitative genetics (statistical designs in which family members are used to estimate genetic influence) and molecular genetics (research designs where genetic influence is estimated by directly measured genotypes). For decades quantitative geneticists used family, twin and adoption designs to estimate genetic and environmental influences on many psychological and psychiatric traits. The strength of these designs is that they can estimate the bottom line of genetic influence regardless of how many genes are involved or how complex their effects, unlike molecular genetics. All quantitative genetic designs have limitations, but each design has different limitations, and, most importantly, their results converge. For example, the adoption design, which typically compares adopted children with their biological and adoptive parents, is limited by possible prenatal factors and selective placement. In contrast, the twin design, which compares monozygotic (MZ) twins who are genetically

*Emerging Trends in the Social and Behavioral Sciences*. Edited by Robert Scott and Stephen Kosslyn.
 ISBN 978-1-118-90077-2.

100% identical, and dizygotic (DZ) twins who share on average 50% of their segregating alleles, is limited by the assumption of equal environments for the two types of twins. These limitations are specific to each design and yet adoption and twin studies generally produce similar estimates of genetic and environmental influences, especially for cognitive traits (Plomin, DeFries, Knopik, & Neiderhiser, 2013). Adoption studies are less common today than twin studies because of the sharp reduction in neonatal adoption, but twin studies continue to proliferate.

Univariate twin studies have demonstrated consistently that virtually every human trait is influenced in part by genetic variation. Multivariate models go beyond this rudimentary finding about the relative influence of nature and nurture on a single phenotype, to investigate the common genetic architecture underlying multiple complex traits. Multivariate twin designs decompose the covariance between multiple phenotypes to provide insights into the common genetic and environmental relationship underlying different traits, as well as the genetic and environmental contributions to their development over time (trait stability and change). Multivariate genetic research has yielded three major findings consistent across many complex traits: heritability increases across development, genetic effects are developmentally stable—strong age-to-age genetic correlations—and are highly pleiotropic, that is, one gene influences many traits. Genetic pleiotropy is inferred from consistent findings of strong genetic covariance between different traits at one measurement occasion: for example, this is true for anxiety and depression (e.g., Thapar & McGuffin, 1997), and for IQ and learning abilities (Davis, Haworth, & Plomin, 2009). Genetic stability is inferred from high genetic correlations between the "same" traits across time. Curiously, despite the observed genetic stability from age to age, many traits appear to show increases in heritability across development (Bergen, Gardner, & Kendler, 2007). Although genetic stability and increasing heritability might appear paradoxical, there are several mechanisms that might explain this. The phenomenon could be driven by a correlation between genes and environments where, for example, children choose environments that suit their genetic propensities (Plomin, DeFries, *et al*., 2013). Imagine a young girl who has a propensity for reading; as she grows up, she chooses environments in which she can be exposed to reading as much as possible. This "tailored" selection will increase the importance of these in explaining individual differences in reading and yet they will be the same genes. It is also possible that heritability increases as new genes come "online" over the lifetime explaining more of the variance (Kendler, Gardner, & Lichtenstein, 2008). Unfortunately, twin designs cannot easily tease out which of these mechanisms are responsible for the phenomena. The most important advance in this field that will help to resolve this issue

and many others will come when specific genetic variants are identified that account for these genetic effects.

## LOW-HANGING FRUIT: MONOGENIC DISORDERS

Mendel, in his original work with peas, was fortunate in that he chose monogenic traits in which a single mutation is necessary and sufficient for the trait to be expressed. For such single-gene traits the underlying genetic variants travel across generations according to predictions based on Mendel's first law of segregation, and they also assort independently from variants of other traits according to Mendel's second law. Unfortunately, complex traits appeared not to conform to these laws, prompting some researchers in the early twentieth century to argue that Mendel's findings were limited to peas, and not "higher-order" species. Although the dispute between the two factions was at times fierce, it was resolved when it became apparent that in complex traits Mendel's laws hold for segregation of alleles at a single locus but are masked by the additive influence of many such loci. The overall genetic influence is driven by many genes, called *polygenic,* and thus the effects of individual genes is small and their overall effect on the phenotype is complex and normally distributed in a population. This is the cornerstone of quantitative genetics expounded by Fisher (1918), who described how Mendel's model could be extended to multiple genes in order to account for inheritance of complex traits.

In molecular genetics, Mendel's law of segregation culminated in huge success in discovering genetic associations with monogenic disorders mainly through implementation of linkage designs using just a few hundred DNA markers across the genome. Linkage analysis uses samples drawn from multigenerational families with affected and unaffected members, where DNA of all family members is "tagged" with several markers whose location on chromosomes is known. The aim is to detect a "link" between one of these markers and a trait or disorder. Linkage successfully detected genetic associations with many Mendelian (monogenic) traits because a mutation in a single locus explained all the variance. However, monogenic disorders affect a tiny proportion of clinical populations; most of the burden of illness involves complex disorders influenced by many genes with small effect sizes. Although linkage has been very successful in identifying genes for single-gene disorders, the method does not have sufficient power to detect genes of small effect size. A technological advance that greatly improved attempts to find genes responsible for heritability of complex traits is genome-wide association (GWA), which is based on allelic association between DNA markers and a trait in unrelated individuals in the population rather than linkage within families.

## FINDING MANY NEEDLES IN THE HAYSTACK: GENOME-WIDE ASSOCIATION STUDIES

The discovery that many complex traits are highly heritable motivated researchers to search for the genetic variants associated with these traits. The pre-GWA era relied mainly on candidate gene association studies, where a specific region of genome was hypothesized to be important for development of a particular trait. Unfortunately, a problem with the candidate gene approach is that the hypotheses generated about the regions we think should be important are often wrong. Many published candidate gene association results is now known to be false positive results (Tabor, Risch, & Myers, 2002). This is particularly telling when replication fails in GWA analysis as this design is hypothesis free in the sense that it does not "care" which region is associated with which trait. An example of such failure is a GWA study of nearly 10,000 individuals where none of the 10 most frequently reported candidate regions for general cognitive ability replicated (Chabris *et al*., 2012).

The genomic era began in the past decade when two major developments provided cost-effective, thorough coverage of the genome. The first event was the completion of the Human Genome Project (H.G.P., 2001), which resulted in the first detailed maps of the human genome and the patterns of linkage disequilibrium (the nonrandom association of alleles across neighboring genetic loci) for hundreds of thousands SNPs (single-nucleotide polymorphism—variation in a single genomic base-pair). The second event was the production of DNA arrays or "chips." A single DNA chip could genotypes hundreds of thousands of SNPs. For each SNP, the DNA chip includes many probes, which is a short DNA sequence containing the SNP. Understanding linkage disequilibrium patterns across the genome was crucial because it showed that careful selection of a few hundred thousand DNA markers was sufficient to comprehensively tag the genome's 3 billion DNA base pairs for GWA studies. DNA variants close together on a chromosome violate Mendel's second law by segregating together producing high between-SNP correlations. A single marker from DNA variants in high linkage disequilibrium is sufficient to "tag" the region. This knowledge was used to select SNPs for DNA chips. In addition, coverage of the genome could be improved by imputing millions of additional SNPs from existing reference maps, such as HapMap (Gibbs *et al*., 2003) or, more recently, the 1000 Genomes Project (Siva, 2008) without actually genotyping these additional SNPs. This meant that it was now possible to interrogate the whole genome simultaneously for association between allele frequencies of individual SNPs and any trait varying in unrelated individuals. This

method is known today as a *GWA* (e.g., Balding, 2006; Cardon & Bell, 2001; Hirschhorn & Daly, 2005).

In GWA, associations are usually examined one SNP at the time, assuming an additive model, although other models can also be implemented. Since current genetic chips can assess millions of SNPs, the correction for multiple testing is daunting, and the number of possible false positives is large; the accepted *p*-value threshold for genome-wide significance is $p < 5 \times 10^{-8}$, which can be thought of as $p < 0.05$ with a Bonferroni correction for one million statistical tests. Despite such stringent correction, within the three years between 2005 and 2008, more than 400 SNP associations for just over 120 traits (www.genome.gov/GWAstudies) achieved this stringent genome-wide threshold, and revealed unexpected insights about genetic influences on complex traits (Visscher, Brown, McCarthy, & Yang, 2012). Arguably the most useful biological insights in genetic associations have been mainly limited to medical traits. For example, in Crohn's disease, many SNPs reported through GWA studies were found in and around genes involved with autophagy, the cell's maintenance process that breaks down dysfunctional components of the cell (WTCCC, 2007). In addition, the same study showed that type 2 diabetes was associated with loci encoding for proteins relevant to insulin secretion, and not insulin signaling, as previously thought. Nevertheless, even if functionally not so obvious, many new loci have also been identified in anthropometric (e.g., Speliotes *et al*., 2010; Visscher, 2008) and some psychiatric traits (e.g., Gershon, Alliey-Rodriguez, & Liu, 2011; Purcell *et al*., 2014). Unfortunately successes in cognitive, social, and behavioral complex traits are yet to come.

Despite GWA successes, fewer associations were reported for complex traits than expected, and associations that were identified could together only account for a small fraction of twin-estimated heritability. The phenomenon is known as "missing heritability," which is discussed in the next section. The important lesson from GWA research is that the largest effect sizes of individual loci of common SNPs are incredibly small and require sample sizes in the hundreds of thousands to achieve genome-wide significance. This realization led to the sudden emergence and proliferation of world-wide consortia, which paid off by doubling the GWA hits to just under 9000 SNPs in more than 700 traits (www.genome.gov/GWAstudies/). Nonetheless, the yield was still not as substantial as expected in that the largest effect sizes, such as the association between the FTO gene and body mass index (BMI), are less than 1% of the variance (Speliotes *et al*., 2010), which implies that the smallest effect sizes are likely to be infinitesimal. This means that it will be difficult to detect and replicate associations with complex traits in the social and behavioral sciences. However, once several

genes are found that are associated with these traits, their effects can be aggregated as polygenic scores, as described in the following section.

## TOGETHER WE ARE STRONGER: POLYGENIC SCORES

Even though the effects of individual SNPs are very small, their effects can be aggregated to increase the total amount of variance explained and thus increase power. For example, if we had 10 individual SNPs, each with an effect size of 0.5%, a polygenic score comprising all of these would account for 5% of the variance, which means that a sample of 150 individuals would have 80% power to detect their cumulative effect. For this reason, polygenic scores are now being used in a new generation of hypothesis-free candidate gene association studies and are the way that the DNA revolution will come to the social and behavioral sciences.

The creation of a polygenic score takes into account the direction of the effect (i.e., which allele is the "increasing" allele) and it can also weight the associations by the magnitude of their effect. To give priority to stronger associations, the scores can be weighted by the betas from the regression (i.e., their effect size). The associations can then be summed, similar to summing items on a scale. Although summing scores is most commonly used, non-additivity can also be incorporated. In addition, SNPs can either be selected from previously reported significant "hits" (polygenic scores—PGS) or they can be amassed by selecting all associations below an arbitrary *p*-value from GWA studies, a genome-wide PGS (Wray *et al*., 2014). The former method is preferred because of the trade-off between specificity and size but the choice depends on the availability of previously reported robust "hits." The tiny effect sizes and moderate-to-large estimates of heritability suggest that the more SNPs you "pool" together the more variance you should explain. However, adding a large number of SNPs with no effect or effects in the opposite direction can attenuate the signal.

Application of polygenic predictors to medical disorders had some success due at least in part to availability of financial resources and large sample sizes. It has been shown that PGS of 150 SNPs accounted for 5% of the variance in the liability for coronary artery disease (Deloukas *et al*., 2013). For bipolar disorder, SNPs accounted for between 1% and 3% of the liability (Psychiatric Gwas Consortium Bipolar Disorder Working & Group, 2011) and a PGS of significant SNPs from discovery GWAS in schizophrenia, accounted for approximately 1% of the liability in the independent sample (Schizophrenia Psychiatric Genome-Wide Association Study, 2011). In the schizophrenia study, extending the PGS to genome-wide PGS increased variance explained to almost 6%. PGS research on quantitative traits has largely been limited to weight and height, and other complex continuous

traits have had to resort to genome-wide PGS. GWA meta-analysis of BMI revealed that 32 replicated SNPs accounted for only ~1.5% of the variance in BMI in independent samples, but a genome-wide PGS that included thousands of SNPs explained ~5% of the variance (Speliotes *et al*., 2010). For height, 180 SNPs accounted for 10% of the variance, and genome-wide PGS increased the variance explained to 13% (Lango Allen *et al*., 2010). Importantly, investigation of behavioral traits of direct interest to social and cognitive scientists are now achieving sample sizes that enable more powerful genomic interrogation. Application of genome-wide PGS to these traits is producing results that are similar to medical and anthropometric traits (height and weight). Specifically, adding more SNPs increases amount of variance explained up to a point. For example, an increase of variance explained with an addition of more SNPs was reported for total years of education (1% using 3500 SNPs and 2.5% using 2.5 million SNPs; Rietveld, Medland, *et al*., 2013) and childhood IQ (using polygenic predictor at $p < 6 \times 10^{-5}$ explained 0.5–1.2% increasing to 3.5% at $p < 0.001$; Benyamin *et al*., 2014). These findings are important despite the small proportions of variance explained, because they suggest that complex traits are indeed highly polygenic.

It is now obvious that even though inclusion of a large number of SNPs consistently increases the amount of variance explained, the gap between twin-estimated heritability and these PGS estimates is still very wide. The phenomenon known today as the "missing heritability" continues to pose many questions, which are as yet unanswered. Twin studies suggest that almost every human trait from our biology through cognition, behavior and even the environment is heritable, with genetic influences usually explaining moderate (30–40%) to high (80–90%) proportions of variance. GWA studies have accounted for only a small fraction of this heritability. The most general explanation is that the influence of each individual SNP is so small that most of the GWA studies thus far have been greatly underpowered to detect them. Another likely source of missing heritability is rarer variants—the markers selected for the DNA arrays were limited to common variants only (minor allele frequency>1%). This meant that many potential "true" associations with SNPs of lower allele frequency could be missed owing to low linkage disequilibrium with the markers. Another possible source of missing heritability is nonadditive effects, such as gene–environment or gene–gene interactions because GWA is limited to additive genetic effects. Finally, it is possible that heritability estimates derived from twin data are inflated. To answer some of these questions, a new quantitative genetic technique has emerged, called genome-wide complex trait analysis (GCTA), as described in the following section.

## HERITABILITY WITHOUT TWINS: GENOME-WIDE COMPLEX TRAIT ANALYSIS (GCTA)

A recently developed method uses genome-wide genotyping data in a novel way to address these questions. The linear mixed model implemented in the GCTA package estimates the amount of phenotypic variance (Yang, Lee, Goddard, & Visscher, 2011) and co-variance (Lee, Yang, Goddard, Visscher, & Wray, 2012) that can be explained by additive effects of all common SNPs tagged by DNA arrays. Several other methods have been developed with the same intent (Wray *et al*., 2014) but GCTA is currently most widely used. GCTA itself has been given different names, such as linear mixed model and genomic-relationship-matrix restricted maximum likelihood. However, clever usage of the four DNA base letters makes *GCTA* the catchiest acronym, and this is what we call it henceforth.

GCTA uses genome-wide genotype data from unrelated individuals to estimate genetic influence on a trait that can be explained by all SNPs included on DNA chips. Because GCTA is based on genome-wide DNA data alone, it can be used to estimate genetic influence for unrelated individuals rather than requiring special relatives such as MZ and DZ twins. For that reason, it bypasses some of the assumptions of the twin method, although it has its own set of assumptions (Plomin, Haworth, Meaburn, Price, & Davis, 2013). GCTA can only detect genetic influence owing to the additive effects of common SNPs that are included on currently available DNA arrays and cannot evaluate the contribution of any specific DNA locus. Nonetheless, GCTA provides important information about the extent to which the genetic architecture of complex traits includes additive effects of common SNPs, and sets the limit for detecting associations in GWA studies.

GCTA analyses have shown that information captured by current DNA arrays can explain a substantial amount of the variance in complex traits, including human height (Yang *et al*., 2010), BMI (Llewellyn, Trzaskowski, Plomin, & Wardle, 2013), psychiatric and medical disorders (Lee, Wray, Goddard, & Visscher, 2011; Lee *et al*., 2012; Lubke *et al*., 2012), cognitive traits (Deary *et al*., 2012; Plomin, Haworth, *et al*., 2013), and economic and political preferences (Benjamin *et al*., 2012). All of these GTCA heritability estimates are approximately half of twin-estimated heritability. In contrast, initial research in psychopathology and personality is less consistent, showing near zero variance explained for most analyses. For example, analyses of a wide range of behavioral problems (symptoms of anxiety and depression, hyperactivity, conduct problems) show negligible SNP heritability despite moderate to high twin heritability estimated within the same sample (Trzaskowski, Dale, & Plomin, 2013; Trzaskowski, Eley, *et al*., 2013). SNP heritabilities of neuroticism and extraversion were reported as

0.06 (SE 0.03) and 0.12 (SE 0.03) respectively (Vinkhuyzen *et al*., 2012), and similar near-zero GCTA estimates have also been shown for other aspects of personality (Verweij *et al*., 2012), including wellbeing (Rietveld, Cesarini, *et al*., 2013). The results are puzzling as all of the studies are adequately powered to detect the expected effect sizes (~40% of twin heritability) and are not limited to self-report questionnaire. For example, one study examined behavior problems reported by parents and teachers as well as the children themselves; further discussion on this topic can also be found in the same publication (Trzaskowski, Dale, *et al*., 2013).

As was the case for early twin studies, GCTA was initially applied in univariate analyses of one trait at a time, but like twin studies, the model was extended to the bivariate analysis of covariance between traits or across age. For example, bivariate GCTA was first applied to the remarkable phenotypic stability of IQ across 60 years from childhood to later life (phenotypic correlation = 0.63) and suggested that the stability is largely due to genetic stability (genetic correlation = 0.62) (Deary *et al*., 2012). Twin studies also suggested that the genetic stability was present despite heritability of IQ increasing across development. Bivariate GCTA supported this finding of genetic stability despite increasing heritability for IQ (Trzaskowski, Yang, Visscher, & Plomin, 2013) and similar twin and GCTA results were reported for BMI (Llewellyn, Trzaskowski, Plomin, & Wardle, 2014). The most likely explanation for these phenomena is gene–environment correlation (Plomin, DeFries, *et al*., 2013). The high genetic correlation suggests that the same genes influence individual differences in IQ across time. However, as we grow older we increasingly select environments that "match" our genetic predispositions making genetic influence stronger even though the same genes are involved.

Although twin studies have reported for decades that most environments are nearly as heritable as behaviors, this work has been limited to twin-specific environments. GCTA opens up the possibility of investigating genetic influence on family-, neighborhood-, or even country-wide environmental measures that cannot be studied using the twin design because they are shared in common by members of a twin pair (Trzaskowski *et al*., 2014). This feature of GCTA should be particularly interesting to social and behavioral scientists as it emphasizes the important interplay between genes and environments. It shows that environments are not simply "out there" that happen randomly to us, but that our genes shape our experiences through our selection, modification and creation of our environments.

Another widely reported finding from twin analyses is the strong genetic correlation (pleiotropy) across different aspects of cognition and across diverse cognitive abilities (Davis *et al*., 2009; Kovas, Haworth, Dale, & Plomin, 2007; Plomin, DeFries, *et al*., 2013). GCTA studies reported point estimates for genetic correlations highly similar to those reported by twin

studies (Trzaskowski, Davis, *et al*., 2013). GCTA estimates of genetic correlation are similar to twin study estimates of genetic correlation, not only within traditional domains such as verbal and nonverbal (Trzaskowski, Shakeshaft, & Plomin, 2013) but also between intelligence and education-related skills such as reading and mathematics (Trzaskowski, Davis, *et al*., 2013). We can expect many more multivariate GCTA analyses in domains other than cognition, most notably psychopathology where it increasingly appears that there is a great deal of overlap among diverse symptoms (Caspi *et al*., 2013).

In summary, GCTA has shed some light on the nature of the discrepancy between heritability estimates from twin studies and combined variance explained from SNPs identified through GWA. Some of the "missing heritability" is hiding in tiny influences of many common SNPs. We called that gap between GWA and GCTA "missing GWA heritability," because GCTA represents the ceiling for GWA results. This gap can in theory be filled with the additive effects of variants tagged by the SNPs on current DNA chips if samples are large enough to detect nearly infinitesimal effect sizes. GCTA results also suggest that even if the "missing GWA heritability" were filled, there would still be a substantial chunk of genetic influence not accounted for by additive effects of common SNPs. This part of the "missing heritability" could be called "missing GCTA heritability," which falls short of twin study heritability because GCTA only reflect additive effects of common SNPs. We delineated these two parts to emphasize a distinction between missing heritability that is caused by lack of power (the effect sizes of already captured common variants are too small to detect with current sample sizes), and missing heritability that is due to genetic variants that have not yet been captured (e.g., rare variants or nonadditive genetic influences). The gap of missing GWA heritability can be narrowed by the brute force of larger samples, but how can DNA variants responsible for missing GCTA heritability be identified?

## 3 Billion Base Pairs: Whole-Genome Sequencing

Whole-genome sequencing involves genotyping all 3 billion base pairs of DNA, rather than just a million or so SNPs as on current genetic chips. Whole-genome sequencing is the end of the story of genetic variation in the sense that all we inherit from our parents is differences in DNA base pair sequences. Since current genetic chips only capture common SNP, whole-genome sequencing could identify more genetic loci associated with complex traits because it captures variants of any kind, not just common SNPs. Rare variants may play a particularly important role in the extreme tails of a trait's distribution. For example, common SNPs influence BMI across the distribution of the normal population, but very rare Mendelian

mutations may account for as much as 5% of extreme obesity (Farooqi & O'Rahilly, 2006).

Other clinical phenotypes, such as intellectual disability, schizophrenia, bipolar disorder, and autism, have been associated with an increased burden of rare variants, including idiosyncratic mutations that first occur in that individual and are not inherited from the individual's parents, called *de novo* mutations (Marioni *et al*., 2014; Neale *et al*., 2011; Purcell *et al*., 2014; Stankiewicz & Lupski, 2010). Perhaps rare variants are responsible for a puzzle in findings concerning mild and severe intellectual disability. One study found that mild intellectual impairment was familial, but severe impairment was not (Nichols, 1984). That is, siblings of severely mentally retarded children showed no mental impairment, whereas siblings of children with "mild" mental disability did, suggesting that severe impairment is not genetically related to common variation in mental ability. In general, common (less severe) disorders are likely to be the quantitative extremes of normal variation (Plomin, Haworth, & Davis, 2009), whereas extreme levels of disability could be mainly a result of accumulation of much rarer variants, including *de novo* mutations as well as environmental "mutations" such as perinatal trauma.

Sequencing will also give us richer information about noncoding regions. Not so long ago noncoding regions of the genome were thought of as an evolutionary "junk," but now these noncoding regions are known to play an important role in regulation of genetic expression and even creation of new genes (Mercer, Dinger, & Mattick, 2009; Muotri, Marchetto, Coufal, & Gage, 2007; Shimoni *et al*., 2007). Sequencing these regions will illuminate regulatory networks and thus contribute to our understanding of genetic responses to changes in the environment.

Thus far the expense of whole-genome sequencing has slowed its progress. However, as costs continue to fall, the availability of sequence data will increase exponentially. It has been predicted by Francis Collins, former director of the Human Genome Project and currently director of the US National Institutes of Health, in his excellent book that: "I am almost certain that complete genome sequencing will become part of newborn screening in the next few years .... It is likely that within a few decades people will look back on our current circumstances with a sense of disbelief that we screened for so few conditions" (Collins, 2010). In fact, parents have already begun paying for sequencing their children's DNA (Rochman, 2012). If this prediction is correct, future social and behavioral research, will see a completely different world of data; a world where budgeting for DNA, genotyping or sequencing will be a thing of the past. Genomic data on nearly everyone will be widely available from centralized sources. For this

reason, it is crucial that social and behavioral scientists in the future are able to capitalize on this opportunity to add genomics to their research.

## REFERENCES

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, *7*(10), 781–791. doi:10.1038/nrg1916

Benjamin, D. J., Cesarini, D., van der Loos, M. J., Dawes, C. T., Koellinger, P. D., Magnusson, P. K., … , Visscher, P. M. (2012). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences of the United States of America, 109*(21), 8026–8031. doi:10.1073/pnas.1120666109

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M. J., … , Visscher, P. M. (2014). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry, 19*(2), 253–258. doi:10.1038/mp.2012.184

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics*, *10*(3), 423–433. doi:10.1375/twin.10.3.423

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics*, *2*(2), 91–99. doi:10.1038/35052543

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S. … Moffitt, T. E. (2013). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2*(2), 119–137. doi:10.1177/2167702613497473

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M. … Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science, 23*(11), 1314–1323. doi:10.1177/0956797611435528

Collins, F. (2010). *The language of life: DNA and the revolution in personalised medicine*. London, England: Profile Books.

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009). Learning abilities and disabilities: Generalist genes in early adolescence. *Cognitive Neuropsychiatry*, *14*, 312–331. doi:10.1080/13546800902797106

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D. … Corley, J. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482*(7384), 212–215.

Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R. … Samani, N. J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics, 45*(1), 25–33. doi:10.1038/ng.2480

Farooqi, I. S., & O'Rahilly, S. (2006). Genetics of obesity in humans. *Endocrine Reviews*, *27*(7), 710–718.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Gershon, E. S., Alliey-Rodriguez, N., & Liu, C. (2011). After GWAS: Searching for genetic risk for schizophrenia and bipolar disorder. *American Journal of Psychiatry*, *168*(3), 253–256. doi:10.1176/appi.ajp.2010.10091340

Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H. … Shen, Y. (2003). The international HapMap project. *Nature, 426*(6968), 789–796.

H.G.P. (2001). The human genome. *Nature*, *409*(6822), 745–964.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*(2), 95–108.

Kendler, K. S., Gardner, C. O., & Lichtenstein, P. (2008). A developmental twin study of symptoms of anxiety and depression: Evidence for genetic innovation and attenuation. *Psychological Medicine*, *38*(11), 1567–1575. doi:10.1017/S003329170800384X

Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years 7912. *Monographs of the Society for Research in Child Development*, *72*, 1–144.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F. … Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature, 467*(7317), 832–838. http://www.nature.com/nature/journal/v467/n7317/abs/nature09410.html - supplementary-information

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *American Journal of Human Genetics*, *88*(3), 294–305. doi:10.1016/j.ajhg.2011.02.002

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using SNP-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. doi:10.1093/bioinformatics/bts474

Llewellyn, C. H., Trzaskowski, M., Plomin, R., & Wardle, J. (2013). Finding the missing heritability in pediatric obesity: The contribution of genome-wide complex trait analysis. *International Journal of Obesity*, *37*, 1506–1509.

Llewellyn, C. H., Trzaskowski, M., Plomin, R., & Wardle, J. (2014). From modelling to measurement: Developmental trends in genetic influence on adiposity in childhood. *Obesity (Silver Spring)*, *22*, 1756–1761.

Lubke, G. H., Hottenga, J. J., Walters, R., Laurin, C., de Geus, E. J., Willemsen, G. … Boomsma, D. I. (2012). Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biological Psychiatry, 72*(8), 707–709. doi:10.1016/j.biopsych.2012.03.011

Marioni, R. E., Penke, L., Davies, G., Huffman, J. E., Hayward, C., & Deary, I. J. (2014). The total burden of rare, non-synonymous exome genetic variants is not associated with childhood or late-life cognitive ability. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1781), 20140117. doi:10.1098/rspb.2014.0117

Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*, *10*(3), 155–159.

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., & Gage, F. H. (2007). The necessary junk: New functions for transposable elements. *Human Molecular Genetics*, *16*(R2), R159–R167.

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M. … Daly, M. J. (2011). Testing for an unusual distribution of rare variants. *PLoS Genetics, 7*(3), e1001322. 10.1371/journal.pgen.1001322

Nichols, P. L. (1984). Familial mental retardation. *Behavior Genetics*, *14*(3), 161–170.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). *Behavioral genetics* (6 ed.). New York, NY: Worth Publishers.

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*, 872–878.

Plomin, R., Haworth, C. M., Meaburn, E. L., Price, T. S., & Davis, O. S. (2013). Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychological Science*, *24*(4), 562–568. doi:10.1177/0956797612457952

Psychiatric Gwas Consortium Bipolar Disorder Working, & Group (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, *43*(10), 977–983. doi:10.1038/ng.943

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P. … Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature, 506*(7487), 185–190. doi:10.1038/nature12975

Rietveld, C. A., Cesarini, D., Benjamin, D. J., Koellinger, P. D., De Neve, J. E., Tiemeier, H. … Bartels, M. (2013). Molecular genetics and subjective well-being. *Proceedings of the National Academy of Sciences of the United States of America, 110*(24), 9692–9697. doi:10.1073/pnas.1222171110

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W. … Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340*(6139), 1467–1471. doi:10.1126/science.1235488

Rochman, B. (2012). Want to know my future. Parents grapple with delving into their kids' DNA. *Time*, *180*, 26.

Schizophrenia Psychiatric Genome-Wide Association Study, C (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–976. doi:10.1038/ng.940

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., & Margalit, H. (2007). Regulation of gene expression by small non-coding RNAs: A quantitative view. *Molecular Systems Biology*, *3*(1).

Siva, N. (2008). 1000 Genomes project. *Nature Biotechnology*, *26*(3), 256–256.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U. … Loos, R. J. F. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics, 42*(11), 937–948. http://www.nature.com/ng/journal/v42/n11/abs/ng.686.html-supplementary-information

Stankiewicz, P., & Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annual Review of Medicine*, *61*, 437–455. doi:10.1146/annurev-med-100708-204735

Tabor, H. K., Risch, N. J., & Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nature Reviews Genetics*, *3*(5), 391–397.

Thapar, A., & McGuffin, P. (1997). Anxiety and depressive symptoms in childhood—A genetic study of comorbidity. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*(6), 651–656.

Trzaskowski, M., Dale, P. S., & Plomin, R. (2013). No genetic influence for childhood behavior problems from DNA analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(10), 1048–1056 e1043. doi:10.1016/j.jaac.2013.07.016

Trzaskowski, M., Davis, O. P., DeFries, J., Yang, J., Visscher, P., & Plomin, R. (2013). DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behavior Genetics*, *1–7*. doi:10.1007/s10519-013-9594-x

Trzaskowski, M., Eley, T. C., Davis, O. S. P., Doherty, S. J., Hanscombe, K. B., Meaburn, E. L. … Plomin, R. (2013). First genome-wide association study on anxiety-related behaviours in childhood. *PloS One, 8*(4), e58676. 10.1371/journal.pone.0058676

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A. … Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence, 42*, 83–88. doi:10.1016/j.intell.2013.11.002

Trzaskowski, M., Shakeshaft, N. G., & Plomin, R. (2013). Intelligence indexes generalist genes for cognitive abilities. *Intelligence*, *41*(5), 560–565. doi:10.1016/j.intell.2013.07.011

Trzaskowski, M., Yang, J., Visscher, P. M., & Plomin, R. (2013). *DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12* (Vol. 19, pp. 380–384). *Molecular Psychiatry*.

Verweij, K. J., Vinkhuyzen, A. A., Benyamin, B., Lynskey, M. T., Quaye, L., Agrawal, A. … Medland, S. E. (2012). The genetic aetiology of cannabis use initiation: A meta-analysis of genome-wide association studies and a SNP-based heritability estimation. *Addiction Biology*. 10.1111/j.1369-1600.2012.00478.x

Vinkhuyzen, A. A., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K., Iacono, W. G. … Wray, N. R. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational Psychiatry, 2*, e102. 10.1038/tp.2012.27

Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics*, *40*(5), 489–490. doi:10.1038/ng0508-489

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. doi:10.1016/j.ajhg.2011.11.029

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A. E., Dudbridge, F., & Middeldorp, C. M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry*, *55*(10), 1068–1087. doi:10.1111/jcpp.12295

WTCCC (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678. doi:10.1038/nature05911

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S. D., Henders, A. K., Nyholt, D. R. … Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. http://www.nature.com/ng/journal/v42/n7/suppinfo/ng.608_S1.html

Yang, J., Lee, S., Goddard, M., & Visscher, P. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. doi:10.1016/j.ajhg.2010.11.011

## MACIEJ TRZASKOWSKI SHORT BIOGRAPHY

While studying for his PhD, **Dr Maciej Trzaskowski** has learnt and applied an extensive selection of statistical methods for quantitative (structural equation modeling) and molecular genetics (e.g., GWAS, and GREML), using various programming languages (e.g., R, Python, Unix, and Shell Scripts). Dr. Trzaskowski has a strong methodological propensity and high interest in exploring new methods. He has been awarded Gottesman-Shields Prize for the best PhD thesis 2013 and since the completion of his PhD, Dr. Trzaskowski remained at the Social, Genetic and Developmental Psychiatry Centre as a postdoctoral research worker. To date, he has published 23 papers, 11 of these as first/joint first author. He has built his own network of collaborations with researchers in the Broad Institute, the Queensland Institute of Medical Research, Queensland Brain Institute (the University of Queensland), Vrije Universiteit (VU) Amsterdam, Harvard, amongst others. His research is increasingly recognized (as evidenced by the Behavior Genetics Association Thompson and Fulker Awards received in 2013 and 2014, respectively) and, most recently, by a successful award of a fellowship from British Academy. He has given invited talks at the London School of Economics (by Lord Richard Layard, Emeritus Professor of Economics) and at the fourth annual meeting by the Social Science Genetic Association Consortium (by Professor Phillip Koellinger).

## ROBERT PLOMIN SHORT BIOGRAPHY

Since 1994, **Professor Robert Plomin** has been MRC Research Professor of Behavioral Genetics at the Institute of Psychiatry, King's College London. In 1994, he cofounded and subsequently directed the MRC Social, Genetic and Developmental Psychiatry Centre, whose goal is to bring together genetic and environmental strategies to study behavioral development. In 1995, he launched the Twins Early Development Study (TEDs) of all

twins born in England and Wales in 1994–1996, which focuses on developmental problems in cognition and behavior, and which Professor Plomin continues to direct. He has published more than 500 papers and more than a dozen books, including the major textbook in the field ("*Behavioral Genetics*," Worth Publishers, 6th edition, 2013). His most recent book, coauthored with Kathryn Asbury, focuses on genetics and education ("*G is for Genes: the Impact of Genetics on Education and Achievement*," Wiley Blackwell, 2013). He has received lifetime research achievement awards from the three major international associations in his field. For details, see https://kclpure.kcl.ac.uk/portal/robert.plomin.html.

## RELATED ESSAYS

Telomeres *(Psychology)*, Nancy Adler and Aoife O'Donovan
Social Epigenetics: Incorporating Epigenetic Effects as Social Cause and Consequence *(Sociology)*, Douglas L. Anderton and Kathleen F. Arcaro
Biology and Culture *(Psychology)*, Robert Peter Hobson
Genetics and the Life Course *(Sociology)*, Evan Charney
Genetic and Environmental Approaches to Political Science *(Political Science)*, Zoltán Fazekas and Peter K. Hatemi
Genetics and Social Behavior *(Anthropology)*, Henry Harpending and Gregory Cochran
A Gene-Environment Approach to Understanding Youth Antisocial Behavior *(Psychology)*, Rebecca Waller *et al*.
Behavioral Heterochrony *(Anthropology)*, Victoria Wobber and Brian Hare