Frequently Asked Questions (FAQs)

This document provides information about the study:

Karlsson Linnér *et al.* 2019. "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics*.

The document was prepared by several of the study's coauthors and draws from and builds on the FAQs for earlier SSGAC papers. It has the following sections:

1. **Background**
2. **Study design and results**
3. **Social and ethical implications of the study**
4. **Appendices**

For clarifications or additional questions, please contact Jonathan P. Beauchamp (jonathan.pierre.beauchamp@gmail.com).

Table of Contents

1 Background

1.1 Who conducted this study? What is the group's overarching goal?

The authors are members of the Social Science Genetic Association Consortium (SSGAC). The SSGAC is a multi-institutional, multi-disciplinary, international research group that aims to identify statistically robust links between genetic variants (for instance, base-pairs of DNA that vary across people) and phenotypes of interest to social scientists. A "phenotype" refers to anything that may be influenced by DNA, such as disease risk or physical characteristics. The phenotypes of interest to social scientists include behaviors, preferences, personality traits, and socioeconomic outcomes.

The SSGAC was formed in 2011 to overcome a specific set of scientific challenges. As is now well understood (Chabris et al. 2015), most phenotypes—including virtually all social-science phenotypes—are influenced by hundreds or thousands of genetic variants. Although in combination their collective effects can be sizeable, almost every one of these genetic variants has an extremely small effect on its own. To reliably identify these individual variants, therefore, scientists must study large samples; typically, hundreds of thousands of individuals are required. One approach to obtaining a large enough sample is for many research groups to pool analyses of their data into a single, large study. This approach has borne considerable fruit when used by medical geneticists interested in a range of diseases and conditions (Visscher et al. 2017a). Most of these advances would not have been possible without large research collaborations between multiple research groups interested in similar questions. The SSGAC was formed in an attempt by social scientists to adopt this research model.

The SSGAC is organized as a working group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), a successful medical consortium. It was founded by three social scientists—Daniel Benjamin (University of Southern California), David Cesarini (New York University), and Philipp Koellinger (Vrije Universiteit Amsterdam)—who believed that genetic data could have a substantial positive impact on research in the social sciences, and that social-science genetics could make important contributions to medical research. The Advisory Board for the SSGAC is composed of prominent researchers representing various disciplines: Dalton Conley (Sociology, New York University), George Davey Smith (Epidemiology, University of Bristol), Tõnu Esko (Molecular Genetics, Broad Institute and Estonian Genome Center), Albert Hofman (Epidemiology, Harvard University), Robert Krueger (Psychology, University of Minnesota), David Laibson (Economics, Harvard University), James Lee (Psychology, University of Minnesota), Sarah Medland (Statistical Genetics, QIMR Berghofer Medical Research Institute), Michelle Meyer (Bioethics, Geisinger Health System), and Peter Visscher (Statistical Genetics, University of Queensland).

The SSGAC is committed to the principles of reproducibility and transparency. Prior to conducting genetic association studies, power calculations are carried out to determine the necessary sample size for the analysis (assuming realistically small effect sizes associated with individual genetic variants). These, together with an analysis plan, are often preregistered on the Open Science Framework (OSF)[a]. Major SSGAC publications are usually accompanied by a FAQ document (such as this one). The FAQ document is written to communicate to journalists and the public what was found and what can and cannot be concluded from the research findings.

[a] The analysis plan for this study can be downloaded here: https://osf.io/cjx9m/.

The SSGAC's first major project was a genome-wide association study (GWAS) of educational attainment published in *Science* (Rietveld et al. 2013b). The study is summarized in a FAQ posted on the SSGAC website (https://www.thessgac.org/faqs). The study was followed by two related studies, using successively much larger samples, published in *Nature* (Okbay et al. 2016b) and *Nature Genetics* (Lee et al. 2018). Subsequent SSGAC papers have studied subjective well-being, depressive symptoms, the personality trait neuroticism, cognitive performance, and reproductive behavior. These papers have been published in *Nature Genetics* (Barban et al. 2016, Okbay et al. 2016a), *Proceedings of the National Academy of Sciences* (Rietveld et al. 2013a, 2014b), and *Psychological Science* (Chabris et al. 2012, Rietveld et al. 2014a), among other journals. The present study is the SSGAC's first study that focuses on the genetics of general risk tolerance.

1.2 The current study focuses on a variable called "general risk tolerance." What is general risk tolerance?

Risk pervades many aspects of human life and is a central concept in the study of decision-making and behavior. Somewhat surprisingly, then, there is no universally agreed-upon definition of "risk." For our purposes, we define "risk" as the degree of variability in possible outcomes, and "risk tolerance" as a person's willingness to choose options that entail more risk, typically to have the chance of obtaining a more rewarding outcome. For example, an engineer with a high degree of risk tolerance would be more willing to quit her job at a stable, large corporation and join a risky start-up. An individual with a high degree of risk tolerance may also be more likely to drive faster than the speed limit on a highway, thus incurring a higher risk of having an accident or a traffic ticket in order to save time.

An individual's risk tolerance typically varies across domains of behavior. For instance, an individual may be willing to take relatively more risks in the career and financial domains, but not in the health and leisure domains. Nonetheless, individuals with greater risk tolerance in one domain are statistically more likely to exhibit greater risk tolerance in other domains as well. For this reason, survey-based measures of *general* risk tolerance—defined as a person's general willingness to take risks—have been used as all-around predictors of risky behaviors such as portfolio allocation, occupational choice, smoking, drinking alcohol, and starting one's own business (Beauchamp et al. 2017, Dohmen et al. 2011, Falk et al. 2015). In our study, we analyze a measure of general risk tolerance based on responses to questions such as: "Would you describe yourself as someone who takes risks? Yes / No." The exact phrasing and number of response categories varied across the study cohorts, but all questions asked subjects about their overall or general attitudes toward risk.

1.3 What was already known about the genetics of risk tolerance prior to this study?

Researchers have found that identical twins (who share all of their genes) tend to be more similar to one another in terms of their risk tolerance than fraternal twins (who share, on average, only half of their genes), which suggests that genetic factors influence risk tolerance. With some assumptions, it is possible to translate the greater similarity of identical twins into an estimate of the "heritability" of risk tolerance. The heritability of risk tolerance is the percentage of the variation in risk tolerance among individuals that can be accounted for statistically by genetic differences, given current environmental conditions. Estimates from twin studies suggest that risk

tolerance is moderately heritable (~30%) (Beauchamp et al. 2017, Cesarini et al. 2009, Harden et al. 2017). We note, however, that such estimates are based on several assumptions and vary across studies, in part because different studies use different measures of risk tolerance as well as different assumptions and methods.

As we further discuss in FAQ 2.2, the current study also estimated the "SNP heritability" of risk tolerance, which is the percentage of the variation in risk tolerance among individuals that can be accounted for statistically by "common SNPs" (a type of genetic variants; see FAQ 2.1 for details), given current environmental conditions. Our estimate suggests that common SNPs account for only ~5% to 9% of the variation in risk tolerance across individuals. Importantly, while these heritability estimates all suggest that genetic factors influence risk tolerance, we emphasize that this does not imply that risk tolerance is pre-determined at birth or that genetic factors act independently of the environment, as we discuss below in FAQs 3.2 and 3.4.

Risk tolerance has been one of the most studied phenotypes in social science genetics. To date, however, nearly all published studies attempting to discover the genetic variants associated with risk tolerance have been "candidate-gene studies" conducted in relatively small samples, ranging from a few hundred to a few thousand individuals. A candidate-gene study tests the associations between a phenotype of interest and a few selected genetic variants that are hypothesized to be associated with the phenotype. Though there is nothing wrong in principle with such studies, we now know that the sample sizes of the candidate-gene studies for risk tolerance and other behavioral traits were probably too small to robustly identify genetic variants[b] (Chabris et al. 2012, Hewitt 2012). Indeed, as we explain in FAQ 2.6, we used our own results to assess the evidence in favor of the main biological pathways and genetic variants which previous candidate-gene studies had hypothesized or reported to relate to risk tolerance. Although our sample was several orders of magnitude larger than the samples used in the candidate-gene studies, we found no evidence that these biological pathways and genetic variants are associated with risk tolerance.

To the best of our knowledge, prior to our study there had only been two studies with samples that were large enough to provide sufficient statistical power to robustly detect genetic variants with small effect sizes (Day et al. 2016, Strawbridge et al. 2018). From these studies, only two genetic variants associated with risk tolerance had been identified.

In summary, when our study was initiated, despite much interest, little was known about which genetic variants are related to risk tolerance.

2 Study design and results

2.1 What did you do in this paper? How was the study designed?

We performed the largest-to-date genome-wide association study (GWAS) of risk tolerance. In a GWAS, scientists look across the human genome for genetic variants that are associated with a phenotype of interest. If a genetic variant is associated, then individuals who have a certain "allele"

[b] As mentioned above, it is now well established that the bulk of the genetic variation in the vast majority of behavioral phenotypes is attributable to a large number of genetic variants, each having a very small effect (Chabris et al. 2015). For that reason, large samples are needed to detect individual genetic variants.

(i.e., a certain version of that variant) are more likely than those with a different allele to exhibit a phenotype (in this case, higher general risk tolerance).

We chose a GWAS study design because it has been a successful research strategy for identifying genetic variants associated with many traits and diseases, including body height (Wood et al. 2014), BMI (Locke et al. 2015), Alzheimer's disease (Lambert et al. 2013), and schizophrenia (Ripke et al. 2014). GWAS have also recently been used to identify genetic variants associated with a variety of health-relevant social science outcomes, such as the number of children a person has (Barban et al. 2016), happiness (Okbay et al. 2016a, Turley et al. 2018), and educational attainment (Okbay et al. 2016b, Rietveld et al. 2013b). Furthermore, scientists who have attempted to replicate reported GWAS associations in independent samples of sufficiently large size have typically been successful (Visscher et al. 2017b), thereby indicating that GWAS associations are robust findings.

In our GWAS of general risk tolerance, we tested ~9.3M single nucleotide polymorphisms (SNPs) from across the human genome for association with general risk tolerance. SNPs are the most common type of genetic variant in the genome and are the genetic variants that are captured by the genetic data used in our study and most other modern genome-wide association studies. (There are other types of genetic variants, which we did not analyze.) Some SNPs have alleles that are relatively common in the population and are called "common SNPs," while other SNPs have one allele that is rare in the population; our GWAS analyzed both common SNPs and some rare SNPs.

As mentioned above, genetic variants associated with social-science phenotypes tend to have very small individual effects on the phenotypes. Therefore, in order to have sufficient statistical power to discover SNPs associated with risk tolerance, we pooled the results from analyses of two very large datasets, the UK Biobank (n = 431,126 individuals) and a dataset of research participants from 23andMe (n = 508,782 individuals), thereby yielding a "discovery" sample of 939,908 individuals. We replicated the findings from this discovery sample in a "replication" sample comprised of ten smaller datasets and totaling 35,445 individuals. In all of these samples, to avoid the statistical confounding that arises from studying ethnically diverse populations, we restricted our GWAS to individuals of European ancestries. (For a somewhat technical explanation, see Appendix 1.)

We used the results of our GWAS of general risk tolerance for a wide range of additional analyses. For example, to examine the extent to which SNPs that are associated with risk tolerance also tend to be associated with other phenotypes, we estimated "genetic correlations" between risk tolerance and a wide range of phenotypes (see FAQ 2.3). In addition, in several samples of genotyped individuals, we used individuals' SNP data and the results of our GWAS to construct "polygenic scores" that partially predict individuals' risk tolerance based on their SNP data (see FAQ 2.4). We also performed a suite of bioinformatics analyses to get insight into the biology of risk tolerance (see FAQs 2.5 and 2.6).

In addition to our GWAS of general risk tolerance, we conducted six supplementary GWAS, of six phenotypes related to risk tolerance and risk-taking behaviors. We conducted a GWAS of "adventurousness," defined as the self-reported tendency to be adventurous vs. cautious. We also conducted GWAS of four risky behaviors that each plausibly capture risk taking in a different domain of behavior: "automobile speeding propensity" (the tendency to drive faster than the speed limit), "drinks per week" (the average number of alcoholic drinks consumed per week), "ever smoker" (whether one has smoked more than once or twice), and "number of sexual partners" (the

lifetime number of sexual partners). Finally, we conducted a GWAS of the first principal component of the four risky behaviors. (The first principal component is a variable that captures the common variation across the four risky behaviors and can be interpreted as capturing the general tendency to take risks across domains.) Section 1.2 of our article's **Supplementary Information** provides more detail on the definitions of these phenotypes. The analyses of the six supplementary phenotypes were performed in samples ranging from ~315,000 to ~557,000 individuals. These samples were smaller because of more limited data availability for these phenotypes.

2.2 What did you find in the GWAS?

Our main GWAS identified 124 SNPs associated with general risk tolerance in our discovery sample. The 124 SNPs are located in 99 "loci" (a locus is a small region of the genome). As expected, the estimated individual effects of the 124 SNPs are all very small: none of the SNPs explain more than 0.02% of the variation in general risk tolerance across individuals.

We verified that the 124 SNPs identified in our discovery sample also tend to be associated with general risk tolerance in our replication sample. Because the replication sample was not large enough to provide adequate statistical power to replicate the associations of each of the 124 SNPs individually, we performed a "holistic" replication analysis. This analysis compares the overall agreement in estimates for the 124 SNPs across the discovery and the replication GWAS. This holistic replication was successful, indicating that it is highly unlikely that the results from our discovery sample were driven by chance alone.

We also estimated the "SNP heritability" of risk tolerance. The SNP heritability of a phenotype is the share of the variation in the phenotype that is statistically accounted for by common SNPs, given current environmental conditions (see FAQ 1.3). We used several methods to obtain our estimates. With all methods, we used a set of common SNPs—that is, SNPs that have alleles that are relatively common in the population—to estimate the heritability. Because the different methods make different assumptions and because we applied the different methods to slightly different data, the methods yielded different heritability estimates. Our estimates suggest that common SNPs account for ~5% to 9% of the variation in risk tolerance across individuals. (The true heritability of risk tolerance is likely to be somewhat higher, since other genetic variants, such as rare SNPs and structural genetic variants, are likely to also contribute to variation in risk tolerance.)

Our six supplementary GWAS (of the phenotypes related to risk tolerance and risk-taking behaviors) identified a total of 741 associations between a specific SNP and one of the phenotypes. Because of the lack of suitable replication samples, we did not perform replication analyses for the GWAS of these six phenotypes.

2.3 Are the SNPs associated with higher risk tolerance in your study also associated with other phenotypes?

Yes. Of the 124 SNPs we identified as associated with general risk tolerance, we found that 72 are also associated with one or more of the six supplementary phenotypes related to risk tolerance and

risk-taking behaviors.[c] We also identified several regions of the genome that stood out as being associated with general risk tolerance and with all or most of the six supplementary phenotypes. We verified that the effects of the SNPs in these regions are concordant, such that SNPs associated with higher general risk tolerance are also associated with more risky behavior. This suggests that these regions represent shared genetic influences on risk tolerance and risky behaviors (rather than just being genomic hot spots containing SNPs associated with many different phenotypes).

In addition, we estimated the "genetic correlation" between general risk tolerance and various other phenotypes. The genetic correlation between two phenotypes is a measure of the extent to which the SNPs that affect one phenotype also tend to affect the other phenotype. We found that general risk tolerance is moderately to highly genetically correlated with a range of risky behaviors. General risk tolerance is genetically correlated with the six supplementary phenotypes (which capture various types of risky behavior), with estimates of the genetic correlations ranging from 0.25 to 0.83. General risk tolerance is also moderately to highly genetically correlated with a number of additional risky behaviors, including cannabis use and self-employment. Importantly, the direction of the genetic correlations is in the expected direction, with higher risk tolerance being associated with riskier behavior. Moreover, our estimates of the genetic correlations between general risk tolerance and the supplementary risky behaviors are substantially higher than the corresponding phenotypic correlations[d], implying that general risk tolerance is more strongly associated with these risky behaviors at the genetic level than at the non-genetic (environmental) level. The relatively high genetic correlations between general risk tolerance and risky behaviors suggests the existence of a genetically-influenced "general factor of risk tolerance" that captures a general tendency to take risk across domains of behavior.

We also found that risk tolerance is moderately genetically correlated with several personality and neuropsychiatric phenotypes. Of note, the estimated genetic correlations with the personality traits extraversion ($\hat{r}_g$ = 0.51[e]), neuroticism ($\hat{r}_g$ = –0.42), and openness to experience ($\hat{r}_g$ = 0.33) are highly statistically significant and are substantially larger in magnitude than previously reported phenotypic correlations, pointing to shared genetic influences among general risk tolerance and these personality traits. We also found statistically significant and positive genetic correlations between general risk tolerance and the neuropsychiatric phenotypes ADHD, bipolar disorder, and schizophrenia.

2.4 How much of a particular person's risk tolerance can be predicted from the results of this paper?

Although each *individual* SNP has a very small effect, the GWAS estimates of the SNPs' (very small) effects can be combined to create a "polygenic score," an index that takes into account the effects of many SNPs from across the genome. Because a polygenic score aggregates the information from many SNPs, it can predict far more of the variation in risk tolerance among individuals than any single SNP. We found that polygenic scores constructed using the results of

[c] Equivalently, as we write in the abstract of the paper, of the 99 loci referred to above and that contain the 124 SNPs associated with general risk tolerance, 46 also contain one or more SNPs associated with at least one of the six supplementary phenotypes.

[d] Although measurement error partly accounts for the lower phenotypic correlations, the genetic correlations remain considerably higher even after adjustment of the phenotypic correlations for measurement error.

[e] "$\hat{r}_g$" denotes a genetic correlation estimate.

our GWAS of general risk tolerance explain up to ~1.6% of the variation across individuals in general risk tolerance. While 1.6% is far larger than the amount of variation explained by individual SNPs (less than 0.02%, as noted above), it is small in absolute terms. As we explain in FAQ 3.3, such polygenic scores cannot be used to meaningfully predict a particular person's risk tolerance.

The predictive power of the polygenic scores is so small partly because our estimates of the SNPs' effect sizes are relatively imprecise. As the available sample sizes for GWAS get larger, estimates of the SNPs' effect sizes will become more precise, and the scores' explanatory power will rise; in theory, if environmental conditions remain the same, it should be possible one day to construct a polygenic score whose explanatory power is close to the heritability of risk tolerance. For example, a score constructed using the set of common SNPs we used to estimate the ~5% to 9% SNP heritability of risk tolerance (see FAQ 2.2), may ultimately explain ~5% to 9% of the variation in risk tolerance across individuals.

Although the polygenic scores we constructed have too little explanatory power to usefully predict any individual's risk tolerance, they have sufficient explanatory power to be useful in social science studies, which focus on average or aggregated behavior in the population (not individual outcomes). Indeed, with 80% statistical power (the conventional threshold for adequate power), the effect of our polygenic scores can be detected in a study with 500 individuals. Therefore, the polygenic scores provided by our study can be useful in social science studies that have at least 500 participants and in which the participants' genomes have been measured. (Several datasets commonly used in social science research meet these criteria.)

2.5 What do your results tell us about human biology and brain development?

To gain insights into the biological mechanisms through which genetic variation influences general risk tolerance, we conducted a suite of bioinformatics analyses. Our bioinformatics analyses point to the involvement of the neurotransmitters glutamate and GABA, which were heretofore not generally believed to play a role in risk tolerance. Glutamate is the most abundant neurotransmitter in the body and plays an excitatory role (i.e., when one neuron secretes it onto another, the second neuron is more likely in turn to transmit its own signal). GABA, by contrast, is the main inhibitory transmitter. To our knowledge, with the exception of a recent study (Lee et al. 2018) prioritizing a much larger number of pathways, no published large-scale GWAS of cognition, personality, or neuropsychiatric phenotypes has pointed to clear roles *both* for glutamate and GABA. Our results suggest that the balance between excitatory and inhibitory neurotransmission may contribute to variation in general risk tolerance across individuals.

Perhaps unsurprisingly, our bioinformatics analyses point to a role for the brain and the central nervous system in modulating risk tolerance. Specifically, our analyses point to the involvement of some brain regions that have previously been identified in neuroscientific studies on decision-making, including the prefrontal cortex, basal ganglia, and midbrain.

2.6 How do your results relate to previous research on the genetics of risk tolerance

As mentioned above in FAQ 1.3, risk tolerance has been one of the most studied phenotypes in social science genetics. However, almost all previous studies have been "candidate-gene studies" conducted in relatively small samples, whose limitations are now appreciated.

We used the results of our GWAS to revisit this previous research. We reviewed the literature that aimed to link risk tolerance to biological pathways, and identified five main biological pathways that have been previously hypothesized to relate to risk tolerance: the steroid hormone cortisol, the monoamine neurotransmitters dopamine and serotonin, and the steroid sex hormones estrogen and testosterone. We then tested whether these five biological pathways relate to risk tolerance

To understand how we tested these five biological pathways, it is helpful to first define what a gene is. A "gene" is a sequence of DNA in the genome that codes for a molecule that has a biological function. The human genome has roughly 20,000 to 25,000 genes; although genes comprise only about 1% to 2% of human genome, they have important biological functions. Genes, like other parts of the genome, can contain SNPs.

To test the five biological pathways for association with risk tolerance, thus, we first used external databases created by other researchers to identify the genes that are involved, or are likely to be involved, in each of these five pathways. Then, we conducted various bioinformatics analyses that used the results of our GWAS and tested the hypothesis that SNPs located in the genes involved in each of the five pathways tend to be more strongly associated with general risk tolerance than other SNPs. We found no evidence in support of that hypothesis, suggesting that the five pathways are not particularly important contributors to individual variation in risk tolerance.

We also used our GWAS results to examine whether SNPs located within (or highly correlated with) 15 specific genes, which previous candidate-gene studies had tested for association with risk tolerance, are indeed associated with risk tolerance. Our sample was several orders of magnitude larger than the samples used in the previous candidate-gene studies (as mentioned above in FAQ 1.3, these studies were conducted in relatively small samples). Despite this, we found no evidence that these 15 genes are associated with risk tolerance, and failed to replicate the main associations the previous candidate-gene studies had reported. Our results are consistent with other studies that have found that small-sample candidate-gene studies have a poor replication record (Chabris et al. 2012, Hewitt 2012).

We also note that our discovery GWAS replicated the associations between general risk tolerance and the two SNPs that had previously been found to be associated with general risk tolerance in the two previous studies with large samples (Day et al. 2016, Strawbridge et al. 2018; see FAQ 1.3). This is not surprising, however, since those two studies analyzed data from the UK Biobank, and the UK Biobank is one of the two large datasets we included in our discovery GWAS.

In summary, instead of pointing to the main genetic variants and biological pathways that had previously been hypothesized to relate to risk tolerance, our analyses identified 124 SNPs associated with risk tolerance (see FAQ 2.2), and point to the involvement of the neurotransmitters glutamate and GABA and of several brain regions (see FAQ 2.5).

3 Social implications of the study

3.1 Did you find "the gene for" (or "the genes for") risk tolerance?

No. We did find several genes[f] containing SNPs associated with general risk tolerance, but that does not mean that these genes determine general risk tolerance. The genetic factors we identified are involved in a long chain of biological processes that exert an influence on human behavior, and those processes are intricately entwined with the environment.

In summary, our findings conform with the expectation that variation in risk tolerance across individuals is influenced by at least thousands, if not millions, of genetic variants (Chabris et al. 2015).

3.2 Does this study show that an individual's level of risk tolerance is determined and fixed at conception?

No. A large share of the variation in risk tolerance among individuals is determined by environmental factors, and environmental factors may also interact with genetic factors. As mentioned in FAQ 1.3, twin studies have found that part of the variation in risk tolerance across individuals is statistically accounted for by genetic factors. But even if all of the variation in risk tolerance at a certain point in time were accounted for by genetic factors (which is definitely not the case), this would not rule out the possibility of past or future environmental influences on risk tolerance. For instance, even if poor eyesight were perfectly heritable and hence completely determined by genetic factors (it is not), the invention of eye glasses, contact lenses, and laser surgery would all drastically improve a person's poor genetic outlook for clear vision. On the flip side, environmental trauma (e.g., a poke to the eye) could drastically worsen another individual's genetic outlook for clear vision. The lesson of eyesight as a phenotype is that heritability of a phenotype—even 100% heritability—does not imply biological determinism: environmental factors can still in principle influence the phenotype. And again, risk tolerance is *far* from being perfectly heritable.

3.3 Can you use the results in this paper to meaningfully predict a particular person's risk tolerance?

No, the results cannot be used to meaningfully predict either a particular person's general risk tolerance, nor their likelihood of taking any particular risk and engaging in any particular sort of risky behavior. As mentioned in FAQ 2.4, we used the results of our GWAS of general risk tolerance to construct polygenic scores that can explain up to ~1.6% of the variation across individuals in general risk tolerance. That means that ~98.4% of the variation in general risk tolerance is explained by factors other than the polygenic scores.

As we also mentioned in FAQ 2.4, we expect that future, larger GWAS will allow the construction of polygenic scores with higher predictive power. However, the predictive power of such scores would still pale in comparison to some other scientific predictors. For example, professional weather forecasts correctly predict about 95% of the variation in day-to-day temperatures. Weather forecasters are therefore vastly more accurate forecasters than social science geneticists will ever be.

[f] As mentioned in FAQ 2.6, a gene is a sequence of DNA in the genome that codes for a molecule that has a biological function; genes, like other parts of the genome, can contain SNPs.

We also note that, while the polygenic scores we constructed can't usefully predict any individual's risk tolerance, they can be useful in social science studies, which focus on aggregated behavior in the population.

3.4 Can environmental factors modify the effects of the specific SNPs you identified?

It is a plausible hypothesis that environmental factors are both moderators and mediators of genetic influences on risk tolerance. For example, it is conceivable that some SNPs have alleles[g] that tend to make individuals relatively less risk tolerant, but only when the individuals are exposed to certain environments (e.g., when they experience a traumatic episode). (Such environments factors would be said to "moderate" the influence of those SNPs.) It is also conceivable that some SNPs affect risk tolerance indirectly, by influencing individuals' preferences for certain environments (e.g., by influencing their preferences for socializing with quiet, cautious friends), which may in turn affect their risk tolerance. (Such environments would be said to "mediate" the influence of those SNPs.)

We did not perform any statistical tests of "gene-environment interactions" in our study. (Gene-environment interactions refer to the moderation of genetic influences by environmental factors.) One promising approach for future studies that seek to identify gene-environment interactions will be to use our GWAS results to construct polygenic scores of general risk tolerance, and then test whether environmental or demographic variables moderate the association between the polygenic scores and an outcome of interest.

To facilitate such research, we have made the summary results of our GWAS publicly available on the SSGAC's website (www.thessgac.org); interested researchers who have access to datasets with genotypic data can download these results and use them to construct polygenic scores.

3.5 What policy lessons or practical advice do you draw from this study?

None whatsoever. Any practical response—individual or policy-level—to this or similar research would be extremely premature. In this respect, our study is no different from genome-wide association studies (GWAS) of complex medical outcomes. In medical GWAS research, it is well understood that identifying genetic variants that affect disease risk is merely a first step toward understanding the underlying biology of that disease. It is not sufficient to assess risk for any specific individual. It is not appropriate to base policies and practices on such assessments.

3.6 Could this kind of research lead to discrimination against, or stigmatization of, people with specific genetic variants? If so, why conduct this research?

Unfortunately, like a great deal of research—including, for instance, research identifying genetic variants associated with increased cancer risk—the results can be misunderstood and could be misapplied, including by being used to discriminate against individuals with specific genetic variants (e.g., in insurance markets). Nevertheless, for a variety of reasons, we do not think that

[g] As mentioned above, an allele is a certain version of a genetic variant.

the best response to the possibility that useful knowledge might be misused is to refrain from producing the knowledge.

First, even if we believed that some knowledge (and specifically knowledge about genetic influences on risk-taking behavior) should be forbidden, that goal is unattainable. Behavioral genetics research, including studies of the relationships between genes and a variety of social-science phenotypes, including risk tolerance, is already being conducted by many scientists and other individuals around the world and will continue to be conducted. Not all of this work involves the use of appropriate scientific methods or the transparent communication of results. In this context, researchers who are committed to developing, implementing, and spreading best practices for conducting and communicating potentially controversial research, including behavioral genetics research, arguably have an ethical responsibility to participate in the development and dissemination of this body of knowledge—rather than abstain from it because of its sensitive nature.

An important theme in our earlier work has been to point out that most existing studies in social-science genetics that report genetic associations with behavioral traits have serious methodological limitations, fail to replicate, and are likely to have false-positive findings (Beauchamp et al. 2011, Benjamin et al. 2012, Chabris et al. 2012, 2015). This same point was made in an editorial in *Behavior Genetics* (the leading journal for the genetics of behavioral traits), which stated that "it now seems likely that many of the published [behavior genetics] findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge" (Hewitt 2012). Consistent with this, the current study was unable to replicate the results of previous candidate-gene studies of risk tolerance (see FAQ 2.6). One of the most important reasons why earlier work has generated unreliable results is that the sample sizes were far too small, given that the true effects of individual genetic variants on behavioral traits are tiny.

Second, one should not assume that behavioral genetics research carries only the potential to increase stigmatization. For instance, behavioral phenotypes such as general risk tolerance are often assumed to be fully and equally within the control of every individual. That view of these behaviors likely contributes to a lack of sympathy for those who exhibit a self-destructive level of risk-taking and, perhaps, suboptimal support for programs that attempt to reduce such behavior. Our purpose here is not to advocate for or against any particular policy for addressing risk behaviors; rather, we mean only to point out that a finding that genes do have some influence can reduce, rather than increase, stigma of those who exhibit risk-tolerant or even risk-seeking behavior.

Third, behavioral genetics research has the potential to yield other benefits, especially as sample sizes continue to increase. Foregoing this research necessarily entails foregoing these and any other possible benefits, some of which will likely be the result of serendipity rather than being foreseeable. For instance, identifying variants associated with risk tolerance may lead to insights regarding the underlying biological pathways. To take an example from medicine, genetic variants in the *LMTK2* (lemur tyrosine kinase 2) gene have small effects on an individual's predisposition to prostate cancer. Nonetheless, knowing that this gene is involved can point scientists toward studying what the gene does, which may end up teaching us something critical about the pathology of prostate cancer. The effect from modifying a biological pathway, e.g., with a pharmaceutical, is potentially much larger than the effect of the gene itself. Moreover, although we are not quite there yet, when many genetic variants taken together capture ~10% of the variation across individuals in risk tolerance, this amount of predictive power (while still too low to be relevant for individual

predictions) will be useful for controlling for genetic factors when studying the effect of a policy or program on an outcome that is also affected by risk tolerance. For example, when studying a policy intervention that aims to reduce the use of illicit substances that present health risks, controlling for as many factors as possible, including genetic factors associated with risk taking, can help generate more precise estimates of the effectiveness of the policy.

In sum, the potential benefits of this research, when conducted responsibly, seem reasonable in relation to the risks, especially considering that this research is already being conducted, sometimes with lesser attention to both scientific rigor and thoughtful science communication. We thus agree with the U.K. Nuffield Council on Bioethics, which concluded in a report (Nuffield Council on Bioethics 2002, p. 114) that "research in behavioural genetics has the potential to advance our understanding of human behaviour and that the research can therefore be justified," but that "researchers and those who report research have a duty to communicate findings in a responsible manner." In our view, responsible behavioral genetics research includes sound methodology and analysis of data; a commitment to publish all results, including any negative results; and transparent, complete reporting of methodology and findings in publications, presentations, and communications with the media and the public, including particular vigilance regarding what the results do—and do not—show (hence, this FAQ document).

4 Appendices

Appendix 1: Quality control measures

There are many potential pitfalls that can lead to spurious results in genome-wide association studies (GWAS). We took many precautions to guard against these pitfalls.

One potential source of spurious results is incomplete "quality control (QC)" of the genetic data. To avoid this problem, we used state-of-the-art QC protocols from medical genetics research (Winkler et al. 2014). We supplemented these protocols by a more recent protocol from Okbay et al. (2016a), as well as by developing and applying additional, more stringent QC filters.

Another potential source of spurious results is a confound known as "population stratification" (e.g., Hamer & Sirota 2000). To illustrate, suppose we were conducting a GWAS of height. People from Northern Europe are on average taller than people from Southern Europe, and there are also small differences in how often certain genetic variants occur in Northern and Southern Europe. If we combine samples of Northern and Southern Europeans and perform a GWAS that ignores the regions the individuals come from, then we would find genetic associations for these variants. However, those associations would simply reflect the fact that the variants are correlated with a population (Northern or Southern Europe) and may actually have nothing to do with height.

In our study we were extremely careful to avoid population stratification as much as possible. At the outset, we restricted the study to individuals of European ancestries, since population stratification problems are more severe when including individuals of different ancestries in the same sample. As is standard in GWAS of medical outcomes, we controlled for "principal components" of the genetic data in the analysis; these principal components capture the small genetic differences across populations, so controlling for them largely removes the spurious associations arising solely from these small differences.

After taking these steps to minimize population stratification, we conducted several analyses to assess how much population stratification still remained in our data. First, we analyzed data on 17,684 sibling pairs from the Swedish Twin Registry and the UK Biobank. The key idea underlying our test was to examine if differences in genetic variants across siblings are associated with differences in the siblings' risk tolerance. If so, then these associations cannot be the result of population stratification. The reason is that full siblings (from the same two biological parents) share their ancestry entirely, and therefore differences in their genetic variants cannot be due to being from different population groups. Unfortunately, because our sample of siblings is much smaller than our discovery GWAS sample (939,908 individuals), our estimates of the effects of the genetic variants within the sibling pairs are much less precise than those in the GWAS. However, we can test whether the GWAS results are entirely due to population stratification, because if they were, then the sibling estimates would not line up with the GWAS estimates at all. In fact, we found that the within-family estimates are more similar to the GWAS estimates in both sign and magnitude than would be expected by chance. These results imply that our GWAS results are not entirely due to population stratification. A second analysis, known as a "LD score regression intercept" analysis (Bulik-Sullivan et al. 2015), indicated that there is some, but not much, population stratification in our GWAS results.

Appendix 2: Additional reading and references

Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, et al. 2016. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* 48(12):1462–72

Beauchamp JP, Cesarini D, Johannesson M. 2017. The psychometric and empirical properties of measures of risk preferences. *J. Risk Uncertain.* 54(3):203–37

Beauchamp JP, Cesarini D, Johannesson M, van der Loos MJHM, Koellinger PD, et al. 2011. Molecular genetics and economics. *J. Econ. Perspect.* 25(4):57–82

Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, et al. 2012. The promises and pitfalls of genoeconomics. *Annu. Rev. Econom.* 4(1):627–62

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, et al. 2015. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47(3):291–95

Cesarini D, Dawes CT, Johannesson M, Lichtenstein P, Wallace B. 2009. Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* 124(2):809–42

Chabris CF, Hebert BM, Benjamin DJ, Beauchamp JP, Cesarini D, et al. 2012. Most reported genetic associations with general intelligence are probably false positives. *Psychol. Sci.* 23(11):1314–23

Chabris CF, Lee JJ, Cesarini D, Benjamin DJ, Laibson DI. 2015. The fourth law of behavior genetics. *Curr. Dir. Psychol. Sci.* 24(4):304–12

Day FR, Helgason H, Chasman DI, Rose LM, Loh P-R, et al. 2016. Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* 48(6):617–23

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG. 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* 9(3):522–50

Falk A, Dohmen T, Falk A, Huffman D. 2015. The nature and predictive power of preferences: Global evidence. *IZA Discussion Papers.*

Hamer DH, Sirota L. 2000. Beware the chopsticks gene. *Mol. Psychiatry*. 5(1):11–13

Harden KP, Kretsch N, Mann FD, Herzhoff K, Tackett JL, et al. 2017. Beyond dual systems: A genetically-informed, latent factor model of behavioral and self-report measures related to adolescent risk-taking. *Dev. Cogn. Neurosci.* 25:221–34

Hewitt JK. 2012. Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav. Genet.* 42(1):1–2

Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, et al. 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* 45(12):1452–58

Lee J, Wedow R, Okbay A, Kong E, Maghzian O, et al. 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50:1112–21

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 518(7538):197–206

Nuffield Council on Bioethics. 2002. Genetics and human behaviour: the ethical context. Nuffield Council on Bioethics [http://nuffieldbioethics.org/wp-content/uploads/2014/07/Genetics-and-human-behaviour.pdf], London

Okbay A, Baselmans BML, Neve J-E De, Turley P, Nivard MG, et al. 2016a. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* 48(6):624–33

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, et al. 2016b. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 533:539–42

Rietveld CA, Cesarini D, Benjamin DJ, Koellinger PD, De Neve J-E, et al. 2013a. Molecular genetics and subjective well-being. *Proc. Natl. Acad. Sci.* 110(24):9692–97

Rietveld CA, Conley DC, Eriksson N, Esko T, Medland SE, et al. 2014a. Replicability and robustness of GWAS for behavioral traits. *Psychol. Sci.* 25(11):1975–86

Rietveld CA, Esko TT, Davies G, Pers TH, Turley PA, et al. 2014b. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.* 111(38):13790–94

Rietveld CACA, Medland SESE, Derringer J, Yang J, Esko T, et al. 2013b. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science.* 340(6139):1467–71

Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, et al. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511(7510):421–27

Strawbridge RJ, Ward J, Cullen B, Tunbridge EM, Hartz S, et al. 2018. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Transl. Psychiatry*. 8(1):1–11

Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, et al. 2018. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* 50(2):229–37

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, et al. 2017a. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* 101(1):5–22

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, et al. 2017b. 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* 101(1):5–22

Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, et al. 2014. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* 9(5):1192–1212

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46(11):1173–86