FAQs about "Genetic variants associated with subjective well-being, depressive symptoms and neuroticism identified through genome-wide analyses"

This document was prepared by several of the co-authors of the paper and Advisory Board members of the Social Science Genetic Association Consortium. For clarifications or additional questions, please contact: Daniel Benjamin (djbenjam@usc.edu).

1. What is the Social Science Genetic Association Consortium (SSGAC)?

The SSGAC is a research infrastructure designed to stimulate dialogue and cooperation among medical researchers, geneticists, and social scientists. The SSGAC facilitates collaborative research that seeks to identify associations between specific genetic variants (small segments of DNA that differ across people) and social science variables, such as behavior, preferences, personality, well-being, and mental health. One major impetus for the formation of the SSGAC was the growing recognition that with respect to most human traits, even though the *joint* effects of many thousands of genetic variants can be substantial, any *individual* genetic variant has a very weak effect. Consequently, very large samples are required to accurately measure the effect of each particular variant. A decade ago, medical researchers began responding to a similar recognition—that most effects of individual genetic variants on complex diseases are very small—by forming research consortia in which groups collaborate by pooling results from many datasets. These efforts have borne considerable fruit, including recent findings on the genetics of autism (Gaugler et al., 2014), schizophrenia (Ripke et al., 2014), and many other diseases and conditions (Visscher et al., 2012). The SSGAC is an attempt to encourage analogous pooling among social-science geneticists. It is organized as a working group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), a successful medical consortium.

The SSGAC was founded by three social scientists (Daniel Benjamin, David Cesarini, and Philipp Koellinger) who believe that genetic data could have a substantial positive impact on research in the social sciences, yet are troubled by how some work in social-science genetics is conducted and communicated. 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), Robert Krueger (Psychology, University of Minnesota), David Laibson (Economics, Harvard), Sarah Medland (Statistical Genetics, QIMR Berghofer Medical Research Institute), Michelle Meyer (Bioethics, Clarkson University and Icahn School of Medicine at Mount Sinai), 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 posted on the Open Science Framework's preregistration website. In many cases, publications are accompanied by a FAQ document (such as this one). The FAQ document is written to communicate to the public what was and was not found and what can and cannot be concluded from the research findings.

The first major project of the SSGAC was a large-scale genome-wide association study (GWAS) on educational attainment, whose results were published in *Science* (Rietveld et al., 2013). The paper was accompanied by a FAQ document posted on the SSGAC website: http://www.thessgac.org/#!faqs/e0udx. Subsequent work of the SSGAC has been published in (or is in press at) *Nature*, *Proceedings of the National Academy of Sciences, Psychological Science,* and other journals.

2. What is "subjective well-being"?

In a nutshell, subjective well-being is the term that social scientists use to describe human psychological well-being, which is usually self-assessed. More precisely, subjective well-being is a catch-all category that includes many specific ways of measuring psychological well-being. One facet of subjective well-being is positive affect, which refers to the emotions a person is experiencing at a particular moment of time. Typical survey questions to measure positive affect include "During the past week, I was happy" and "How would you rate your emotional wellbeing at present?" Another facet of subjective well-being is life satisfaction, which refers to a longer-term, higher-level evaluation of one's life. A typical survey question would be "How satisfied are you with your life as a whole?" Positive affect and life satisfaction are different from each other but are highly correlated nevertheless.

In our study, we combined different survey measures of positive affect and life satisfaction. This strategy allowed us to assemble much larger samples than prior work and to maximize statistical power to discover genetic associations.

A drawback of our research strategy is that mixing different measures of subjective well-being makes any discovered associations more difficult to interpret. For that reason, research isolating specific, high quality measures of the various facets of subjective well-being (as well as depressive symptoms and neuroticism) is an important next step. Our results will facilitate such work because the genetic variants that we identify can be used as candidate genes for follow-on studies conducted in smaller samples with fine-grained measures of subjective well-being.

3. What do you mean by "depressive symptoms" and "neuroticism"?

The variable we call "depressive symptoms" is closely related to depression. Depression is a psychiatric condition characterized by feelings of sadness, anxiety, low energy, bodily aches and pains, pessimism, and other symptoms. Researchers often study depression by administering questionnaires to ask subjects if they are experiencing the symptoms of depression. The researchers then divide the survey respondents into two groups: those who are depressed and those who are not.

Instead of dividing respondents into two groups (i.e., binary categorization into a depressed group and a non-depressed group), we created a single continuous scale/spectrum that we call "depressive symptoms." All respondents are placed somewhere on this continuous scale, depending on their survey responses. The scale is constructed so that respondents who have more depressive symptoms have a higher scale value. We decided to study depressive symptoms rather than the binary categories "depressed/non-depressed" because using a

continuous scale gives us greater statistical power. Binary categorization throws away information that has statistical value, like symptom variation within each of the binary categories.

Neuroticism is a personality trait characterized by easily experiencing negative emotions such as anxiety and fear. Like other personality traits, it is usually measured by questionnaires that ask people to report about their own personality and behaviors. Here too we constructed a continuous scale that represents the degree of neuroticism.

As in other genetic studies of depression and neuroticism, our analysis combined data from different studies that used different surveys to measure these traits.

4. What was already known about the genetics of subjective well-being, depression, and neuroticism prior to this study?

Twin and family studies have found that genetic differences across individuals can lead to differences in subjective well-being, depression, and neuroticism. Such studies have also found that these three traits share some of the same genetic factors in common.

Although genetic factors in general are known to play a role in these traits, few specific genetic variants have been identified. Our study is the first genome-wide association study (GWAS) of subjective well-being. There have been a few genome-wide association studies of depression (Cai et al., 2015; de Moor et al., 2015; Ripke et al., 2013) and neuroticism (de Moor et al., 2015), but these have found fewer genetic variants, probably because the sample sizes in these studies were relatively small. Concurrently with our study, a GWAS of neuroticism using a subset of our sample reports similar findings to our neuroticism findings (Smith et al., in press).

5. What did you do in this particular study?

Our primary analysis is a genome-wide association study (GWAS) of subjective well-being based on a sample of 298,420 individuals. We were able to obtain this sample size by combining results from separate analyses conducted in 59 different cohorts of individuals. This analysis is one of the largest genome-wide association studies ever conducted for a behavioral trait.

In our study, we also conducted genome-wide association studies of depressive symptoms in a sample of 161,460 individuals and neuroticism in a sample of 170,911 individuals. For these analyses, we combined results from previously published papers (de Moor et al., 2015; Ripke et al., 2013) with new analyses of additional data.

We subsequently partnered with a large, ongoing study of depression (Hyde et al., under review) in a sample of roughly 368,890 individuals who are customers of the personal genomics company 23andMe. We used this new dataset to replicate the genetic associations that we reported for depressive symptoms and neuroticism.

In our analyses, we examined approximately 2.5 million genetic variants called single nucleotide polymorphisms, or SNPs. SNPs are the smallest and most common type of genetic

variant (ways in which the genomes of people can differ), but they are not the only type of genetic variant. Another type of genetic variant is an inversion polymorphism. An inversion polymorphism is a large segment of the genome that is reversed end to end, or inverted, in some people. In our SNP data, we can sometimes statistically detect the presence of an inversion polymorphism. In some of our analyses, we examined inversion polymorphisms in addition to SNPs. Inversion polymorphisms are especially interesting because they tend to have larger effects than SNPs, and far fewer inversion polymorphisms than SNPs have been identified as associated with human traits.

The results of these genetic association analyses are the core scientific contribution of our paper. We conducted several additional analyses to shed some light on possible biological mechanisms underlying our findings and to explore the genetic correlations between the three phenotypes we studied and various health outcomes.

6. What did you find?

In our GWAS of subjective well-being (in our sample of roughly 300,000 individuals), we identified three SNPs.

In our GWAS of depressive symptoms (in our sample of roughly 180,000 individuals), we identified two SNPs.

In our GWAS of neuroticism (in our sample of roughly 170,000 individuals), we identified nine SNPs and two inversion polymorphisms.

In our joint analyses of the three traits, we identified two additional SNPs associated with neuroticism and two associated with both depressive symptoms and neuroticism. We also found that most of the genetic variants associated with depressive symptoms and/or neuroticism are also associated with subjective well-being, and vice-versa.

In our replication sample from an ongoing study of depression (of roughly 370,000 additional individuals), both of the SNPs that we found to be associated with depressive symptoms replicated. We also found that the eleven genetic variants that we found to be associated with neuroticism showed up strongly in the depression replication sample. (We did not study the SNPs that we found to be associated with subjective well-being in the replication sample because some of the individuals in the replication sample were also in the sample for the GWAS of subjective well-being. This sample overlap would have biased the analysis.)

The estimated effect sizes of the genetic variants are small. For subjective well-being, each SNP we identified explains only 0.01% of the variation across individuals. Each of the SNPs associated with depressive symptoms and neuroticism account for only 0.02% to 0.04% of the variation of these outcomes in the population. Since an inversion polymorphism affects much more of the genome than a SNP, we expected that the inversion polymorphisms we identified would have a larger effect size. We were able to estimate the effect size of one of the inversion polymorphisms that we found to be associated with neuroticism. The inversion polymorphism does in fact have a larger effect size—roughly 0.06% of the variation in neuroticism—but this effect size is still small. By way of comparison, the largest effect sizes that have been found for SNPs associated with height and BMI are 0.4% and 0.3% of the

variation, respectively—an order of magnitude larger than those we found for the behavioral traits we study.

Our finding that *individual* genetic variants have very weak associations with these outcomes confirms that very large samples—such as the hundreds of thousands of individuals that we studied—are necessary to accurately detect genetic variants associated with them. Accurately identifying more genetic variants would require larger samples and/or more accurate outcome measures than we had available. Our results support the view that many more genetic variants with depression (and the other traits we study) will be identified when the available sample sizes become even larger (Hyman, 2014).

There is no contradiction between our finding that the effect sizes of individual genetic variants are small and the findings from previous work that a substantial share of the variation across individuals in subjective well-being, depression, and neuroticism can be attributed to genetic factors (e.g., some studies estimate roughly 40%). These findings taken together imply that the genetic influences on these traits result from the cumulative effects of at least thousands (probably millions) of different genetic variants, not just a few.

7. How do we know that the GWAS results are not spurious?

There are many potential pitfalls that can lead to spurious results in genome-wide association studies (GWAS) such as ours. 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).

Another potential source of spurious results is a confound known as "population stratification" (Hamer and Sirota, 2000). To illustrate, suppose we were conducting a GWAS on 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 ignoring the origins of the individuals, 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 employed multiple strategies that reduce the impact of population stratification. At the outset, we restricted the study to individuals of European descent, since population stratification problems are more severe when including European-descent and non-European-descent individuals in the same sample. As is standard in GWAS on 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 a number of analyses to assess how much population stratification still remained in our data. The results of these tests indicate that there is very little stratification in our estimates.

We conducted additional tests to confirm that our GWAS results for subjective well-being are not driven by this remaining population stratification. To do so, we used a subset of the individuals in our data, 4,869 sibling pairs (from three of the datasets that contributed to our study). The key idea underlying our tests is to examine if *differences* in genetic variants across siblings are associated with *differences* in the siblings' subjective well-being. If so, then these associations cannot be the result of population stratification. The reason is that full siblings (from the same genetic parents) share their ancestry entirely, and therefore differences in their genetic variants cannot be due to being from different population groups (in fact, genetic differences between siblings are random). Unfortunately, because our sample of siblings (~9,000 individuals) is much smaller than our overall GWAS sample (~300,000 individuals), our estimates of the effects of the genetic variants within the sibling pairs are much noisier than 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 at all with the GWAS estimates. In fact, we find 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 solely due to population stratification.

The results of a number of the other analyses in the paper provide additional reassurance that our GWAS results are not spurious. For example, the findings from our analyses of genetic overlap between the three traits we focus on and other outcomes are similar to findings from prior studies that examined some of the same outcomes that we do.

8. What did you find in additional analyses?

Our genetic-association results served as a starting point for several additional analyses:

(i) *Identifying the extent of genetic overlap between subjective well-being, depressive symptoms, and neuroticism*. Because we have data on ~9 million SNPs, we can estimate the extent of genetic overlap between these traits far more precisely than prior studies that used the similarity between twins and family members (rather than using direct measurement of genetic data). Specifically, we can estimate the extent of genetic overlap by examining how strongly the SNPs associated with one of the traits are associated with the other traits. We find that the three traits are strongly genetically overlapping, with pairwise genetic correlations of roughly 0.8 in magnitude.

(ii) *Identifying the extent of genetic overlap between our three traits—subjective well-being, depressive symptoms, and neuroticism—and other outcomes*. Using similar methods, we can also estimate the extent of genetic overlap between our three traits and other outcomes that have been studied in GWAS with large samples. We examined five physical health outcomes that are known or believed to be risk factors for poor health: body mass index, ever-smoker status, coronary artery disease, fasting glucose, and triglycerides. We also examined five neuropsychiatric outcomes: Alzheimer's disease, anxiety disorders, autism spectrum disorder, bipolar disorder, and schizophrenia.

We find rather weak genetic overlap with all five of the physical health phenotypes, as well as with Alzheimer's disease and autism spectrum disorder. We find moderate genetic overlap with schizophrenia and bipolar disorder and strong genetic overlap with anxiety disorders. In

fact, the genetic correlations between our three traits and anxiety disorders are of similar magnitude as the genetic correlations of our three traits with each other. This finding suggests that future studies of the genetics of anxiety disorders may benefit by analyzing anxiety disorders jointly with the three traits on which we focus.

(iii) *Investigating biological pathways*. We can draw inferences about biological pathways using methods (from bioinformatics) that synthesize the patterns of association from many SNPs across the genome. In general, these methods examine whether genes known to be involved in particular biological systems are especially likely to be associated with our three traits.

Using such a method, we find that across our three traits, genetic variants regulating gene expression in the central nervous system and adrenal/pancreas tissues are strongly enriched for association. The cause of the adrenal/pancreas enrichment is unclear, but we note that the adrenal glands produce several hormones, including cortisol, epinephrine, and norepinephrine, known to play important roles in the bodily regulation of mood and stress. More speculatively, some of our biological analyses aimed to pinpoint specific genes that are promising candidates for further investigation in relation to the traits we study. One of these genes is *DRD2*, which encodes the D_2 subtype of the dopamine receptor, a target for antipsychotic drugs that is also known to play a key role in neural reward pathways. Another such gene is *MAPT*, which has previously been reported to be involved in neurodegenerative disorders, including Parkinson's disease and progressive supranuclear palsy, a rare disease whose symptoms include depression and apathy.

9. What policy lessons do you draw from this study?

None. *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 known genetic variants are not yet predictive enough of complex diseases to have significant value for assessing the risk to any given individual. Our current paper shows that most genetic effects on the outcomes we studied are even smaller and more diffuse than the genetic associations estimated with typical medical phenotypes.

10. Did you find "the genes" for subjective well-being, depression, and neuroticism?

No. We did not find "the genes" for the outcomes we studied. Characterizing the results this way would be misleading for several reasons.

First, subjective well-being, depression, and neuroticism are primarily determined by environmental factors.

Second, the explanatory power of each individual genetic variant that we identify is extremely small. Our results show that the genetic influences on the outcomes we study are comprised of thousands, or even millions, of genetic variants, each of which matters a little bit.

Third, environmental factors are likely to amplify or attenuate the impact of specific genetic

variants (and may affect which genetic variants are associated with well-being, depression, and neuroticism).

11. Does this study show that an individual's level of subjective well-being (or neuroticism, or risk of being depressed) is determined at birth?

No. This is probably one of the most common misconceptions about genetics research. Even if it were true—and it is certainly not—that genetic factors accounted for *all* of the differences among individuals in subjective well-being, it would *still* not follow that an individual's subjective well-being is "determined" at birth (or, more accurately, at conception). There are at least three reasons for this:

First, some genetic effects may operate *through* environmental channels. As an illustrative example, *suppose* that the genetic variants we identified influence how extraverted, or outgoing, an individual is. Furthermore, *suppose* that being more extraverted helps a person to make more friends, which in turn makes the person happier. In this example, changes to the intermediate environmental channel—number of friends—could have drastic effects on the outcome of happiness. Indeed, the genetic association might not be found at all in environments in which a person's number of friends is less strongly related to extraversion, such as in a close-knit community where everyone knows each other.

Second, even if the genetic effects on well-being operated entirely through non-environmental mechanisms that are difficult to modify (such as direct influences on the neurotransmitters that operate in the brain's reward pathways), there could still exist powerful environmental interventions that, if implemented, would change the genetic relationships. In a famous example suggested by the economist Arthur Goldberger, even if all the variation in unaided eyesight were due to genes, there could still be enormous benefits from introducing eyeglasses. Indeed, the environmental intervention of eyeglasses often counteracts 100% of the effect of genes on eyesight (Goldberger, 1979). Similarly, policies that aim to reduce differences in subjective well-being (e.g., through redistribution that makes society more egalitarian, thereby reducing differences in happiness that result from income inequality) may counteract the effects of genetic predisposition on subjective well-being.

Third, even if the genetic effects on subjective well-being were not altered by changes in the environment, those environmental changes themselves could still have a major impact on the subjective well-being of the population as a whole. For example, if economic progress enabled people to work fewer hours, then everyone might have more leisure, and the population as a whole might become happier. By analogy, 80%–90% of the variation across individuals in height is due to genetic factors. Yet the current generation of people is much taller than past generations—all due to changes in the environment (such as improved nutrition).

References

Cai, N. *et al.* Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523,** 588–591 (2015).

De Moor, V. D. B. et al. Genome-wide association study identifies novel locus for neuroticism and shows polygenic association with Major Depressive Disorder. *JAMA Psychiatry* **72,** 642–650 (2015).

Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat. Genet.* **46,** 881–5 (2014).

Goldberger, A. S. Heritability. *Economica* **46,** 327–347 (1979).

Hamer, D. H., & Sirota, L. Beware the chopsticks gene. *Mol. Psychiatry* **5,** 11–13 (2000).

Hyde, C. L. et al. Common genetic variants associated with major depressive disorder among individuals of European descent. *Nat. Genet.* Under review.

Hyman, S. Mental health: Depression needs large human-genetics studies. *Nature* **515,** 189–191 (2014).

Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–1471 (2013).

Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).

Smith, D. J. *et al.* Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *Molecular Psychiatry*. In press.

Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90,** 7–24 (2012).

Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9,** 1192–1212 (2014).