nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0309-3

# Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences

**Humans vary substantially in their willingness to take risks. In a combined sample of over 1 million individuals, we conducted genome-wide association studies (GWAS) of general risk tolerance, adventurousness, and risky behaviors in the driving, drinking, smoking, and sexual domains. Across all GWAS, we identified hundreds of associated loci, including 99 loci associated with general risk tolerance. We report evidence of substantial shared genetic influences across risk tolerance and the risky behaviors: 46 of the 99 general risk tolerance loci contain a lead SNP for at least one of our other GWAS, and general risk tolerance is genetically correlated ($|\hat{r}_g|$ ~0.25 to 0.50) with a range of risky behaviors. Bioinformatics analyses imply that genes near SNPs associated with general risk tolerance are highly expressed in brain tissues and point to a role for glutamatergic and GABAergic neurotransmission. We found no evidence of enrichment for genes previously hypothesized to relate to risk tolerance.**

Choices in important domains of life, including health, fertility, finance, employment, and social relationships, rarely have consequences that can be anticipated perfectly. The degree of variability in possible outcomes is called risk. Risk tolerance—defined as the willingness to take risks, typically to obtain some reward—varies substantially across humans and has been actively studied in the behavioral and social sciences. An individual's risk tolerance may vary across domains, but survey-based measures of general risk tolerance (for example, 'Would you describe yourself as someone who takes risks?') have been found to be good all-around predictors of risky behaviors such as portfolio allocation, occupational choice, smoking, drinking alcohol, and starting one's own business[1–3].

Twin studies have established that various measures of risk tolerance are moderately heritable ($h^2 \sim 30\%$, although estimates in the literature vary[3–5]). Discovery of specific genetic variants associated with general risk tolerance could provide insights into underlying biological pathways, advance understanding of how genetic influences are amplified and dampened by environmental factors, enable the construction of polygenic scores (indexes of many genetic variants) that can be used as overall measures of genetic influences on individuals, and help distinguish genetic variation associated with general versus domain-specific risk tolerance.

Although risk tolerance has been one of the most studied phenotypes in social science genetics, most claims of positive findings have been based on small-sample candidate gene studies (Supplementary Table 1), whose limitations are now appreciated[6]. To date, only two loci associated with risk tolerance have been identified in GWAS[7,8].

Here, we report results from large-scale GWAS of self-reported general risk tolerance (our primary phenotype) and six supplementary phenotypes: 'adventurousness' (defined as the self-reported tendency to be adventurous versus cautious); four risky behaviors: '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 ever been a smoker), and 'number of sexual partners' (the lifetime number of sexual partners); and the first principal component (PC) of these four risky behaviors, which we interpret as capturing the general tendency to take risks across domains. All seven phenotypes are coded such that higher phenotype values are associated with higher risk tolerance or risk taking. Table 1 lists, for each GWAS, the datasets we analyzed and the GWAS sample sizes.

## Results

**Association analyses.** All seven GWAS were performed in European-ancestry subjects, included controls for the top ten (or more) PCs of the genetic relatedness matrix and for sex and birth year (Supplementary Table 2), and followed procedures described in a prespecified analysis plan (see URLs) and in the Supplementary Note.

In the discovery phase of our GWAS of general risk tolerance ($n = 939{,}908$), we conducted a GWAS by using the UK Biobank (UKB; $n = 431{,}126$) and then performed a sample-size-weighted meta-analysis of those results with GWAS results from a sample of research participants from 23andMe ($n = 508{,}782$). The UKB measure of general risk tolerance is based on the question 'Would you describe yourself as someone who takes risks? Yes/No'. The 23andMe measure is based on a question about overall comfort taking risks, with five response options ranging from 'very comfortable' to 'very uncomfortable'. The genetic correlation[9] between the UKB and 23andMe cohorts ($\hat{r}_g = 0.77$, s.e. = 0.02) is <1 but large enough to justify our approach of pooling the two cohorts (see Section 2 in the Supplementary Note of ref. [10] for a theoretical demonstration of the merits of pooling cohorts despite moderate heterogeneity of phenotype measures).

The quantile–quantile (Q–Q) plot (Supplementary Fig. 1a) from the discovery GWAS exhibited substantial inflation ($\lambda_{GC} = 1.41$). According to the estimated intercept from a linkage disequilibrium (LD) score regression[11], only a small share of this inflation (~5%) in test statistics was due to confounding biases such as cryptic relatedness and population stratification. To account for these biases, we inflated GWAS standard errors by the square root of the LD score regression intercept[11].

A full list of authors and affiliations appears at the end of the paper.

We identified 124 approximately independent SNPs (pairwise $r^2<0.1$) that attained genome-wide significance ($P<5\times10^{-8}$). These 124 'lead SNPs' are listed in Supplementary Table 3 and are shown in Fig. 1a. All had coefficients of determination ($R^2$ values) below 0.02%, and the SNP with the largest per-allele effect was estimated to increase general risk tolerance by ~0.026 standard deviations in our discovery sample (Supplementary Fig. 2). To test whether the lead SNPs' effect sizes were heterogeneous across the 23andMe and UKB cohorts, we generated an omnibus test statistic by summing Cochran's *Q* statistics across all lead SNPs; in agreement with our genetic correlation estimate of less than unity between the two cohorts, we rejected the null hypothesis of homogeneity ($P=4.32\times10^{-5}$; Supplementary Note). To define genomic loci around the lead SNPs, we took the physical regions containing all SNPs in LD (pairwise $r^2>0.6$) with the lead SNPs and merged loci within 250 kb of each other; the 124 lead SNPs were located in 99 such loci (Supplementary Table 3). We supplemented those analyses with a conditional and joint multiple-SNP (COJO) analysis[12], which identified 91 genome-wide-significant 'conditional associations' (Supplementary Table 3).

In the replication phase of our GWAS of general risk tolerance (combined $n=35{,}445$), we meta-analyzed summary statistics from ten smaller cohorts. Additional details on cohort-level phenotype measures are provided in Supplementary Table 4. The cohorts' survey questions differed in terms of their exact wording and number of response categories, but all questions asked subjects about their overall or general attitudes toward risk. The genetic correlation[9] between the discovery and replication GWAS was 0.83 (s.e.=0.13). Of the 124 lead SNPs, 123 were available or well proxied by an available SNP in the replication GWAS results. Out of these 123 SNPs, 94 had a concordant sign ($P=1.7\times10^{-9}$), and 23 were significant at the 5% level in one-tailed *t* tests ($P=4.5\times10^{-8}$) (Supplementary Fig. 3). This empirical replication record closely matched theoretical projections that take into account sampling variation and the winner's curse (Supplementary Note).

In the UKB, we tested and confirmed that a much higher fraction of males (34%) than females (19%) described themselves as risk tolerant on the general risk tolerance measure (*t*-test $P<1\times10^{-100}$; Supplementary Fig. 4), a result consistent with much prior research[13,14]. We used bivariate LD score regression to calculate the genetic correlation between GWAS performed separately in the sample of females and in the sample of males in the UKB. Our estimate ($\hat{r}_g=0.822$, s.e.=0.033) was high enough to justify our approach of pooling males and females in our other analyses to maximize statistical power[10]. Nonetheless, our estimate was significantly smaller than unity, suggesting that the autosomal genetic factors contributing to general risk tolerance, although largely similar across sexes, are not identical.

Our six supplementary GWAS—of adventurousness, the four risky behaviors, and their PC ($n=315{,}894$ to 557,923; Supplementary Tables 4 and 5)—were conducted by using methods comparable to those in the primary GWAS, except that they had no replication phases, and most involved a single large cohort. Supplementary Fig. 1 shows Q–Q plots, and Supplementary Fig. 5 shows Manhattan plots.

Table 1 provides a summary overview of the seven GWAS. We identified a total of 864 'lead associations': the sum total of the 124 general-risk-tolerance lead SNPs together with the 740 lead SNPs from the six supplementary GWAS. (These 864 lead associations were obtained by considering each of our seven phenotypes separately and using the standard genome-wide-significance *P*-value threshold of $5\times10^{-8}$. When we instead considered the seven GWAS jointly and use a Bonferroni-corrected *P*-value threshold of $7.1\times10^{-9}$ ($5\times10^{-8}/7$), we obtained 566 lead associations across the seven GWAS.) Because we did not have the data to conduct replication analyses of the lead associations from the supplementary GWAS, we calculated the maxFDR[15], a theoretical upper bound on the false discovery rate (FDR), for each GWAS. The maxFDR estimates were low across all GWAS (the highest estimate was $1.22\times10^{-3}$, for automobile speeding propensity), thus providing reassurance about the robustness of the lead associations.

Applying our locus definition, we identified a total of 703 'locus associations': the sum total of the 99 general-risk-tolerance loci together with the 604 loci from the supplementary GWAS (Supplementary Note). Pooling the loci corresponding to the 703 locus associations, and merging loci within 250 kb of each other, yielded 444 distinct loci. COJO analyses[12] identified a sum total of 655 conditional associations across all seven GWAS. (When we instead considered the seven GWAS jointly and used a Bonferroni-corrected *P*-value threshold of $7.1\times10^{-9}$ ($5\times10^{-8}/7$), we obtained 464 locus associations and 505 conditional associations across the seven GWAS.) We verified that the results of the COJO analyses were consistent with those from multiple regressions, using individual-level genotype-dosage data from the UKB (Supplementary Note). Supplementary Tables 3, 6, and 7 report the lead SNPs, the genomic loci, and the results of the COJO analyses. Table 1 also shows the SNP heritabilities[16] of the seven phenotypes, calculated from the GWAS results; the SNP heritabilities ranged from ~0.05 (for general risk tolerance) to ~0.16 (for the first PC of the four risky behaviors).

We note that 212 of the 864 lead associations were located within long-range LD regions[17] or candidate inversions (that is, genomic regions highly prone to inversion polymorphisms; Supplementary Note). Of these, only 109 were also conditional associations, and 46 were in loci containing no conditional associations, thus indicating that many lead associations in the long-range LD regions and candidate inversions may tag causal variants that are also tagged by other lead associations. We discuss some of these regions in the next section.

**Genetic overlap.** There was substantial overlap across the results of our GWAS. For example, 46 of the 99 general-risk-tolerance loci contained a lead SNP of at least one of the other GWAS, and 72 of the 124 general-risk-tolerance lead SNPs were in weak LD (pairwise $r^2>0.1$) with a lead SNP of at least one of the other GWAS (including 45 for adventurousness and 49 for at least one of the four risky behaviors or their first PC). To empirically assess whether this overlap could be attributed to chance, we conducted resampling exercises under the null hypothesis that the lead SNPs of our supplementary GWAS were distributed independently of the general-risk-tolerance loci and lead SNPs. We strongly rejected this null hypothesis ($P<0.0001$; Supplementary Note).

Several long-range LD regions, candidate inversions, and LD blocks[18] stood out for being associated both with general risk tolerance and with all or most of the supplementary phenotypes. We tested whether the signs of the lead SNPs located in these regions tended to be concordant across our primary and supplementary GWAS. We strongly rejected the null hypothesis of no concordance ($P<3\times10^{-30}$; Supplementary Note), suggesting that these regions represent shared genetic influences, rather than colocalization of causal SNPs. Figure 1b and Supplementary Fig. 6 show local Manhattan plots for some of these long-range LD regions and candidate inversions. The long-range LD region[17] on chromosome 3 (~83.4 to 86.9 Mb) contained lead SNPs from all seven GWAS as well as the most significant lead SNP from the general-risk-tolerance GWAS, rs993137 ($P=2.14\times10^{-40}$), which is located in the gene *CADM2*. Another long-range LD region, on chromosome 6 (~25.3–33.4 Mb), covers the HLA complex and contained lead SNPs from all GWAS except drinks per week. Three candidate inversions on chromosomes 7 (~124.6–132.7 Mb), 8 (~7.89–11.8 Mb), and 18 (~49.1–55.5 Mb) contained lead SNPs from six, five, and all seven of our GWAS, respectively. Finally, four other LD blocks[18] that did

**Table 1 | GWAS results**

| GWAS | Cohorts analyzed | *n* | Mean $\chi^2$ | LD score intercept (s.e.) | No. lead SNPs | No. loci | No. cond. assoc. | SNP $h^2$ (s.e.) |
|---|---|---|---|---|---|---|---|---|
| General risk tolerance (disc. GWAS) | UKB; 23andMe | 939,908 | 1.85 | 1.04 (0.01) | 124 | 99 | 91 | 0.046 (0.001) |
| General risk tolerance (repl. GWAS) | 10 indep. cohorts | 35,445 | 1.03 | 1.00 (0.07) | 0 | 0 | 0 | - |
| General risk tolerance (disc. + repl.) | UKB; 23andMe; 10 indep. cohorts | 975,353 | 1.87 | 1.04 (0.01) | 132 | 107 | 97 | 0.045 (0.001) |
| Adventurousness | 23andMe | 557,923 | 1.98 | 1.05 (0.01) | 167 | 137 | 126 | 0.098 (0.002) |
| Automobile speeding propensity | UKB | 404,291 | 1.53 | 1.03 (0.01) | 42 | 36 | 33 | 0.079 (0.003) |
| Drinks per week | UKB | 414,343 | 1.61 | 1.03 (0.01) | 85 | 62 | 61 | 0.085 (0.003) |
| Ever smoker | UKB; TAG Consortium[43] | 518,633 | 1.97 | 1.05 (0.01) | 223 | 183 | 172 | 0.109 (0.003) |
| Number of sexual partners | UKB | 370,711 | 1.77 | 1.04 (0.01) | 117 | 97 | 88 | 0.128 (0.003) |
| First PC of the four risky behaviors | UKB | 315,894 | 1.77 | 1.05 (0.01) | 106 | 89 | 84 | 0.156 (0.004) |

The table provides an overview of the GWAS of our primary and supplementary phenotypes. Replication analysis of the lead SNPs' association results in independent cohorts was conducted only for the discovery GWAS of general risk tolerance. *n*, GWAS sample size; mean $\chi^2$, mean GWAS $\chi^2$ statistics across HapMap3 SNPs with MAF greater than 0.01; LD score intercept, estimate of the intercept from a LD score regression[11] using HapMap3 SNPs with MAF greater than 0.01; no. lead SNPs, number of approximately independent (pairwise $r^2$ <0.1) lead SNPs; no. loci, number of associated loci; no. cond. assoc., number of conditional associations in the COJO analysis[12]; SNP $h^2$, SNP heritability estimated with the HESS method[16] using 1000 Genomes phase 3 SNPs with MAF greater than 0.05; disc., discovery; repl., replication; indep., independent.

not overlap known long-range LD or candidate inversion regions each contained lead SNPs from five of our GWAS (including general risk tolerance). Although many of the lead SNPs in these regions were not conditional associations, the above results regarding the numbers of GWAS with lead SNPs in these regions also held when we considered only the conditional associations instead of the lead SNPs in those regions. The two long-range LD regions and the three candidate inversions have been found to be associated with numerous phenotypes, including many cognitive and neuropsychiatric phenotypes[19].

To investigate genetic overlap at the genome-wide level, we estimated genetic correlations with self-reported general risk tolerance by using bivariate LD score regression[9]. (For this and all subsequent analyses involving general risk tolerance, we used the summary statistics from the combined meta-analysis of our discovery and replication GWAS.) The estimated genetic correlations with our six supplementary phenotypes were all positive, larger than ~0.25, and highly significant ($P < 2.3 \times 10^{-30}$; Fig. 2), indicating that SNPs associated with higher general risk tolerance also tend to be associated with riskier behavior. The largest estimated genetic correlations were with adventurousness ($\hat{r}_g = 0.83$, s.e. = 0.01), number of sexual partners (0.52, s.e. = 0.02), automobile speeding propensity (0.45, s.e. = 0.02), and the first PC of the four risky behaviors (0.50, s.e. = 0.02).

Our estimates of the genetic correlations between general risk tolerance and the supplementary risky behaviors were substantially higher than the corresponding phenotypic correlations (Supplementary Tables 8 and 9). Although measurement error partly accounts for the low phenotypic correlations, the genetic correlations remained considerably higher even after adjustment of the phenotypic correlations for measurement error. The comparatively large genetic correlations support the view that a general factor of risk tolerance partly accounts for cross-domain correlation in risky behavior[20,21] and imply that this factor is genetically influenced. The lower phenotypic correlations suggest that environmental factors are more important contributors to domain-specific risky behavior[22,23].

To increase the precision of our estimates of the SNPs' effects on general risk tolerance, we leveraged the high degree of genetic overlap across our phenotypes by conducting multitrait analysis of GWAS (MTAG)[15]. We used as inputs the summary statistics of our GWAS of general risk tolerance, of our first five supplementary GWAS (that is, not including the first PC of the four risky behaviors), and of a previously published GWAS on lifetime cannabis use[24] (Supplementary Note). MTAG increased the number of general-risk-tolerance lead SNPs from 124 to 312 (Supplementary Fig. 7 and Supplementary Table 10).

We also estimated genetic correlations between general risk tolerance and 28 additional phenotypes (Fig. 2 and Supplementary Table 9). These included phenotypes for which we could obtain summary statistics from previous GWAS, as well as five phenotypes for which we conducted new GWAS. The estimated genetic correlations for the personality traits extraversion ($\hat{r}_g = 0.51$, s.e. = 0.03), neuroticism (−0.42, s.e. = 0.04), and openness to experience (0.33, s.e. = 0.03) were significantly distinguishable from zero after Bonferroni correction and were substantially larger in magnitude than previously reported phenotypic correlations[25], thus pointing to shared genetic influences among general risk tolerance and these traits. After Bonferroni correction, we also found significant positive genetic correlations with the neuropsychiatric phenotypes attention deficit hyperactivity disorder (ADHD), bipolar disorder, and schizophrenia. When viewed in light of the genetic correlations we found with some supplementary phenotypes and additional risky behaviors classified as externalizing (for example, substance use, elevated sexual behavior, and fast driving), these results suggest the hypothesis that the overlap with the neuropsychiatric phenotypes is driven by their externalizing component[26].

**Polygenic prediction.** We constructed polygenic scores of general risk tolerance to gauge their potential usefulness in empirical research (Supplementary Note). We used the Add Health, HRS, NTR, STR, UKB-siblings, and Zurich cohorts as validation cohorts (Supplementary Table 5 provides an overview of these cohorts; the UKB-siblings cohort comprised individuals with at least one full sibling in the UKB). For each validation cohort, we constructed the score using summary statistics from a meta-analysis of our discovery and replication GWAS that excluded the cohort (for the UKB-siblings cohort, we reran our UKB GWAS after excluding individuals from that cohort). Our measure of predictive power was the incremental $R^2$ (or pseudo-$R^2$) from adding the score to a

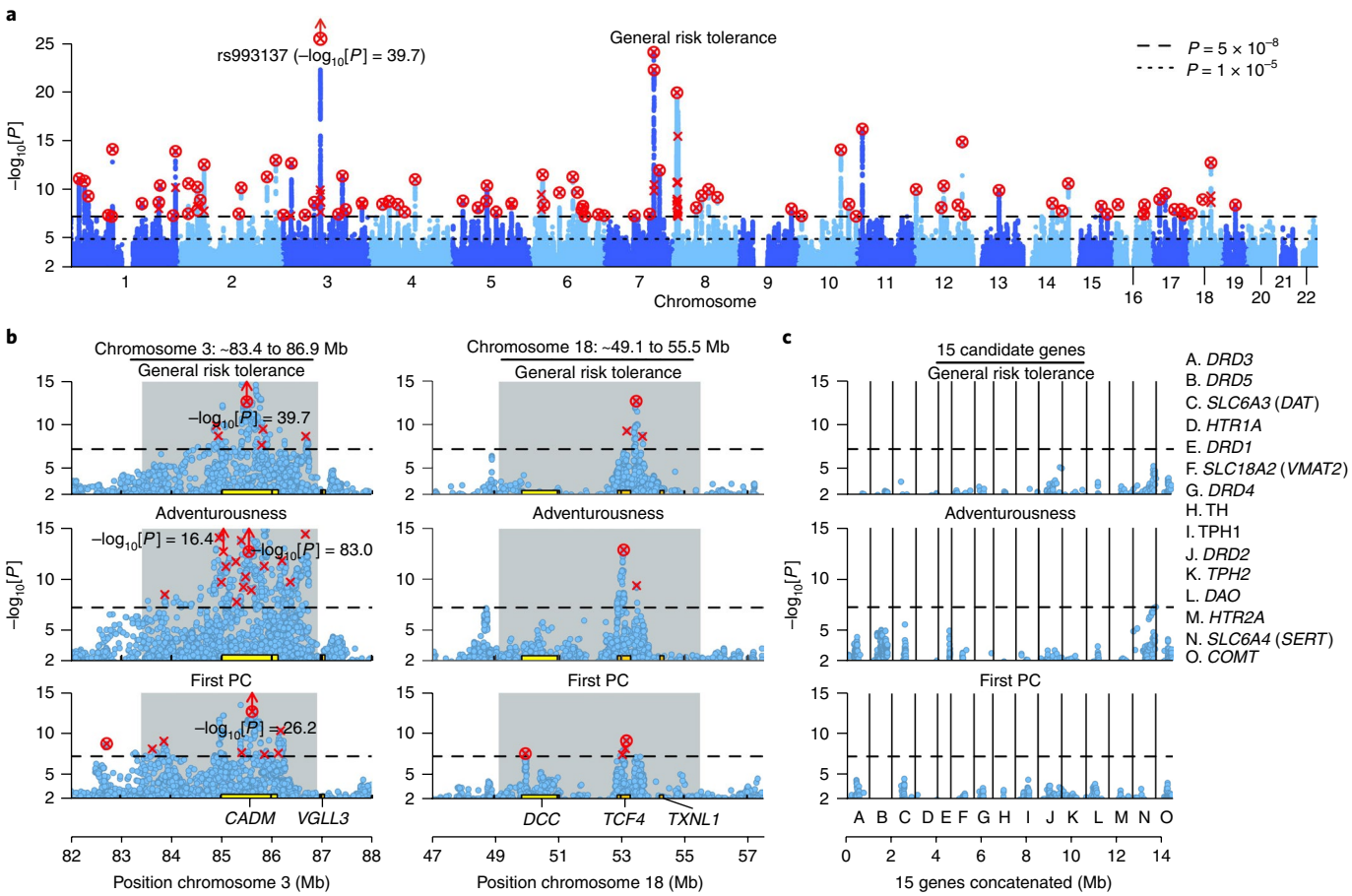


**Fig. 1 | Manhattan plots.** In all panels, the *x* axis is chromosomal position; the *y* axis is the GWAS *P* value on a $-\log_{10}$ scale (based on a two-tailed *z* test); each lead SNP is marked by a red ×; each conditional association is marked by a red ○; and each SNP that is both a lead SNP and a conditional association is marked by a red ⊗. **a**, Manhattan plots for the discovery GWAS of general risk tolerance ($n = 939{,}908$). **b**, Local Manhattan plots of a long-range LD region on chromosome 3 and a candidate inversion on chromosome 18 that contained lead SNPs for all seven of our GWAS. The gray background marks the locations of long-range LD or candidate inversion regions. **c**, Local Manhattan plots of the areas around the 15 most commonly tested candidate genes in the prior literature on the genetics of risk tolerance. Each local plot shows all SNPs within 500 kb of the gene's borders that were in weak LD ($r^2 > 0.1$) with a SNP in the gene. The 15 plots are concatenated and shown together in the panel, divided by black vertical lines. The 15 genes were not particularly strongly associated with general risk tolerance or the risky behaviors, as can be seen by comparing the results within each row across **b** and **c** (the three rows correspond to the GWAS of general risk tolerance, adventurousness ($n = 557{,}923$), and the first PC of the four risky behaviors ($n = 315{,}894$)).

regression of the phenotype on controls for sex, birth year, and the top ten PCs of the genetic relatedness matrix.

Our preferred score was constructed with LDpred[27]. Our largest validation cohort ($n \sim 35{,}000$) was the UKB-siblings cohort. In that validation cohort, the score's predictive power was 1.6% for general risk tolerance, 1.0% for the first PC of the four risky behaviors, 0.8% for number of sexual partners, 0.6% for automobile speeding propensity, and ~0.15% for drinks per week and ever smoker. Across our validation cohorts, in which other phenotypes were measured, the score was also predictive of several personality phenotypes and a suite of real-world measures of risky behaviors in the health, financial, and career domains, and others (Supplementary Figs. 8 and 9 and Supplementary Tables 11–14). The incremental $R^2$ observed for general risk tolerance was consistent with our theoretical prediction, given the GWAS sample sizes, the SNP heritability of general risk tolerance (Table 1), and the imperfect genetic correlations across the GWAS and validation cohorts[28,29] (Supplementary Note).

**Biological annotation.** To gain insights into the biological mechanisms through which genetic variation influences general risk tolerance, we conducted a number of bioinformatics analyses by using the results of the combined meta-analysis of our discovery and replication GWAS of general risk tolerance.

First, we systematically reviewed the literature that aimed to link risk tolerance to biological pathways (Supplementary Note). Our review covered studies based on candidate genes (that is, specific genetic variants used as proxies for biological pathways), pharmacological manipulations, biochemical assays, and genetic manipulations in rodents, as well as other research designs. Our review identified 132 articles that matched our search criteria (Supplementary Table 1). This previous work has focused on five main biological pathways: the steroid hormone cortisol, the monoamines dopamine and serotonin, and the steroid sex hormones estrogen and testosterone. Using a MAGMA[30] competitive gene-set analysis, we found no evidence that SNPs within genes associated with these five pathways tended to be more associated with general risk tolerance than SNPs in other genes (Supplementary Table 15). Furthermore, none of the other bioinformatics analyses reported below point to these pathways.

We also examined the 15 most commonly tested autosomal genes within the dopamine and serotonin pathways, which were the focus of most of the 34 candidate-gene studies identified by our literature review. We verified that the SNPs available in our GWAS results

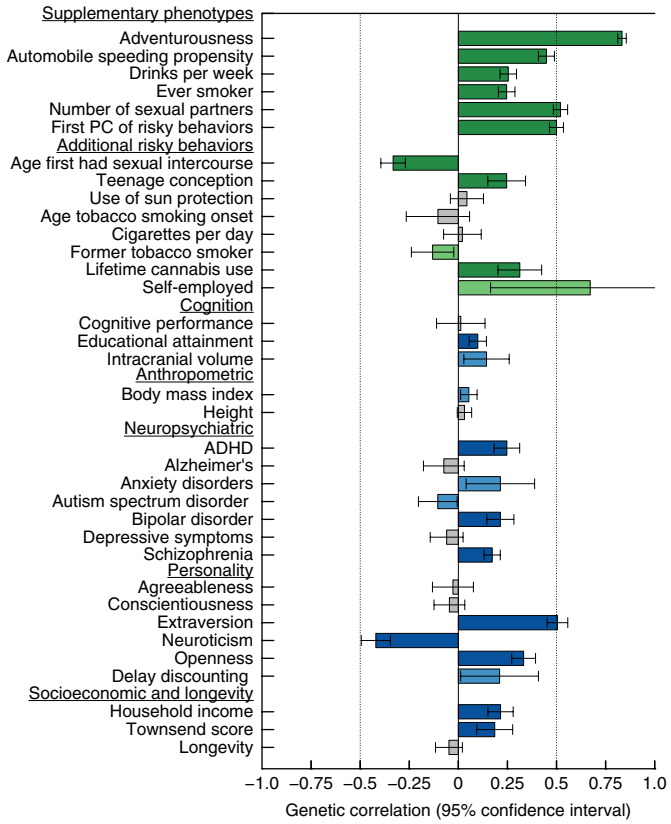


**Fig. 2 | Genetic correlations with general risk tolerance.** The genetic correlations were estimated with bivariate LD score regression[9]. Error bars show 95% confidence intervals. For the supplementary phenotypes and the additional risky behaviors, green bars represent significant estimates with the expected signs, where higher risk tolerance was associated with riskier behavior. For the other phenotypes, blue bars represent significant estimates. Light green and light blue bars represent genetic correlations that were statistically significant at the 5% level, and dark green and dark blue bars represent correlations that were statistically significant after Bonferroni correction for 35 tests (the total number of phenotypes tested). Gray bars represent correlations that were not statistically significant at the 5% level. The two dotted vertical lines indicate genetic correlations of −0.5 and 0.5, respectively. All significance tests are two sided.

tagged most of the genetic variants typically used to test the 15 genes. Across one SNP-based test and two gene-based tests, we found no evidence of non-negligible associations between those genes and general risk tolerance (Fig. 1c and Supplementary Table 16). (We note, however, that some brain regions identified in analyses reported below are areas where dopamine and serotonin play important roles.)

Second, we performed a MAGMA[30] gene analysis to test each of ~18,000 protein-coding genes for association with general risk tolerance (Supplementary Note). After Bonferroni correction, 285 genes were significant (Supplementary Fig. 10 and Supplementary Table 17). To gain insight into the functions and expression patterns of these 285 genes, we looked them up in the Gene Network[31] coexpression database.

Third, to identify relevant biological pathways and identify tissues in which genes near general-risk-tolerance-associated SNPs were expressed, we applied the software tool DEPICT[32] to the SNPs with $P$ values less than $10^{-5}$ in our GWAS of general risk tolerance (Supplementary Note).

Both the Gene Network and the DEPICT analyses separately pointed to a role for glutamate and GABA neurotransmitters, which are the main excitatory and inhibitory neurotransmitters in the brain, respectively[33] (Fig. 3a and Supplementary Tables 18 and 19). To our knowledge, with the exception of a recent study[34] prioritizing a much larger number of genes and pathways, no published large-scale GWAS of cognition, personality, or neuropsychiatric phenotypes has indicated clear roles both for glutamate and GABA (although glutamatergic neurotransmission has been implicated in recent GWAS of schizophrenia[35] and major depression[36]). Our results suggest that the balance between excitatory and inhibitory neurotransmission may contribute to variation in general risk tolerance across individuals.

The Gene Network and the DEPICT tissue enrichment analyses also both separately pointed to enrichment of the prefrontal cortex and the basal ganglia (Fig. 3b and Supplementary Tables 18, 20, and 21). The cortical and subcortical regions highlighted by DEPICT include some of the major components of the cortical-basal ganglia circuit, which is known as the reward system in human and nonhuman primates and is critically involved in learning, motivation, and decision-making, notably under risk and uncertainty[37,38]. We caution, however, that our results do not point exclusively to the reward system.

Finally, we used stratified LD score regression[39] to test for the enrichment of SNPs associated with histone marks in ten tissue or cell types (Supplementary Note). Central nervous system tissues were the most enriched, accounting for 44% (s.e. = 3%) of the heritability while composing only 15% of the SNPs (Supplementary Fig. 11a and Supplementary Table 22). Immune/hematopoietic tissues were also significantly enriched. Although a role for the immune system in modulating risk tolerance is plausible given prior evidence of its involvement in several neuropsychiatric disorders[35,36], future work is needed to confirm this result and to uncover specific pathways that might be involved.

## Discussion

Our results provide insights into biological mechanisms that influence general risk tolerance. Our bioinformatics analyses point to the role of gene expression in brain regions that have been identified by neuroscientific studies on decision-making, notably the prefrontal cortex, basal ganglia, and midbrain, thereby providing convergent evidence with that from neuroscience[37,38]. Yet our analyses failed to find evidence for the main biological pathways that had been previously hypothesized to influence risk tolerance. Instead, our analyses implicate genes involved in glutamatergic and GABAergic neurotransmission, which were heretofore not generally believed to play a noteworthy role in risk tolerance.

Although our focus was on the genetics of general risk tolerance and risky behaviors, environmental and demographic factors accounted for a substantial share of these phenotypes' variation. We observed sizeable effects of sex and age on general risk tolerance in the UKB data (Supplementary Fig. 4), and life experiences have been shown to affect both measured risk tolerance and risky behaviors (see, for example, refs. [40,41]). The GWAS results that we generated should allow researchers to construct and use polygenic scores of general risk tolerance to measure how environmental, demographic, and genetic factors interact with one another.

For the behavioral sciences, our results bear on an ongoing debate about the extent to which risk tolerance is a 'domain-general' as opposed to a 'domain-specific' trait. Low phenotypic correlations in risk tolerance across decision-making domains have been interpreted as supporting the domain-specific view[22,23]. Across the risky behaviors we study, we found that the genetic correlations were considerably higher than the phenotypic correlations (even after the latter were corrected for measurement error) and that many lead SNPs were shared across our phenotypes. These observations suggest that the low phenotypic correlations across domains are due to environmental factors that dilute the effects of a genetically influenced domain-general factor of risk tolerance.

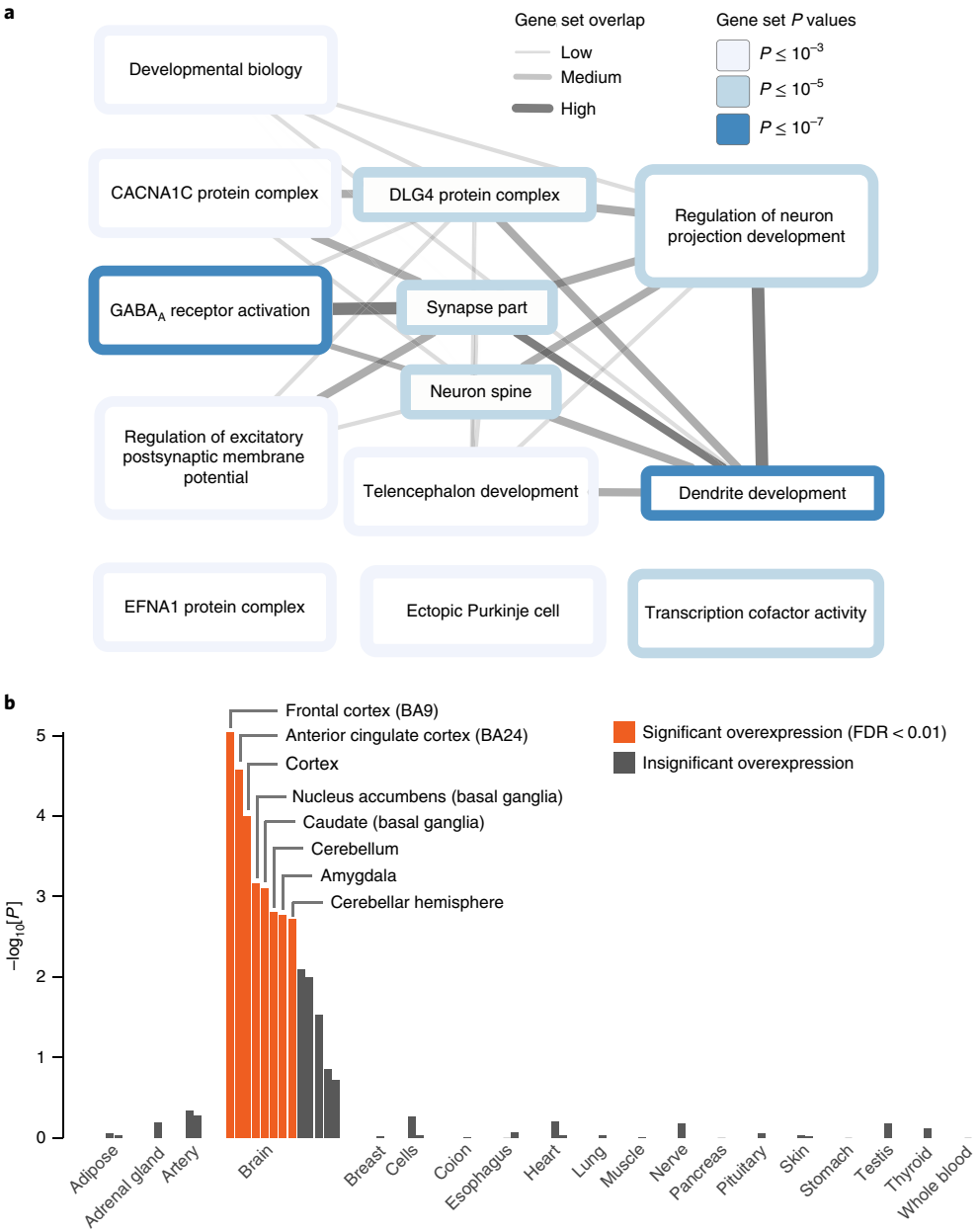


**Fig. 3 | Results from selected biological analyses. a**, DEPICT gene-set enrichment diagram. We identified 93 reconstituted gene sets that were significantly enriched (FDR <0.01) for genes overlapping DEPICT-defined loci associated with general risk tolerance; according to the affinity propagation method[42], these were grouped into the 13 clusters displayed in the graph. Each cluster was named after its exemplary gene set, as chosen by the affinity propagation tool, and each cluster's color represents the permutation *P* value of its most significant gene set. The 'synapse part' cluster includes the gene set 'glutamate receptor activity', and several members of the 'GABA$_A$ receptor activation' cluster are defined by GABA signaling. Overlap between the named representatives of two clusters is represented by an edge. Edge width represents the Pearson correlation $\rho$ between the two respective vectors of gene membership scores ($\rho < 0.3$, no edge; $0.3 \leq \rho < 0.5$, thin edge; $0.5 \leq \rho < 0.7$, intermediate edge; $\rho \geq 0.7$, thick edge). **b**, Results of DEPICT tissue enrichment analysis using GTEx data. The panel shows whether the genes overlapping DEPICT-defined loci associated with general risk tolerance were significantly overexpressed (relative to genes in random sets of loci matched by gene density) in various tissues. Tissues are grouped by organ or tissue type. The orange bars correspond to tissues with significant overexpression (FDR <0.01). The *y* axis is the significance on a $-\log_{10}$ scale. The Supplementary Note provides additional details.

**URLs.** Publicly archived analysis plan for this project, https://osf.io/cjx9m/; Social Science Genetic Association Consortium, https://www.thessgac.org/data; BCFtools, https://samtools.github.io/bcftools/bcftools.html; BEAGLE, http://faculty.washington.edu/browning/beagle/b3.html; BOLT-LMM v.2.3.2, https://data.broadinstitute.org/alkesgroup/BOLT-LMM/; DEPICT (retrieved February 2015), https://data.broadinstitute.org/mpg/depict/; EasyQC v.9.0, http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/; GCTA, http://cnsgenomics.com/software/gcta; HESS, http://bogdan.bioinformatics.ucla.edu/software/hess/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; IMPUTE4, https://jmarchini.org/impute-4/; LD score regression, https://github.com/bulik/ldsc/; LDpred, https://bitbucket.org/bjarni_vilhjalmsson/ldpred; Mach2QTL, http://csg.sph.umich.edu/yli/mach/download/mach2qtl.source.V112.tgz; MAGMA, https://ctg.cncr.nl/software/magma; Minimac2, https://genome.sph.umich.edu/wiki/Minimac2; MTAG software, https://github.com/omeed-maghzian/mtag; PBWT, https://github.com/richarddurbin/

pbwt; PLINK, http://zzz.bwh.harvard.edu/plink/plink2.shtml; Python v.2.7, https://www.python.org/download/releases/2.7/; QCtool v.2, http://www.well.ox.ac.uk/~gav/qctool_v2/; R, https://www.r-project.org/; REGSCAN v.0.2.0, https://www.geenivaramu.ee/en/tools/regscan; Rstudio, https://www.rstudio.com/; ShapeIT, http://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; SMR, https://cnsgenomics.com/software/smr/; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; Stata v.14.2, https://www.stata.com/install-guide/windows/download/

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0309-3

Received: 26 January 2018; Accepted: 7 November 2018;
Published online: 14 January 2019

## References

1. Dohmen, T. et al. Individual risk attitudes: measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* **9**, 522–550 (2011).
2. Falk, A., et al. *The Nature and Predictive Power of Preferences: Global Evidence* (IZA Institute of Labor Economics, 2015).
3. Beauchamp, J. P., Cesarini, D. & Johannesson, M. The psychometric and empirical properties of measures of risk preferences. *J. Risk Uncertain.* **54**, 203–237 (2017).
4. Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P. & Wallace, B. Genetic variation in preferences for giving and risk taking. *Q. J. Econ.* **124**, 809–842 (2009).
5. Harden, K. P. et al. Beyond dual systems: a genetically-informed, latent factor model of behavioral and self-report measures related to adolescent risk-taking. *Dev. Cogn. Neurosci.* **25**, 221–234 (2017).
6. Hewitt, J. K. Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav. Genet.* **42**, 1–2 (2012).
7. Day, F. R. et al. Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* **48**, 617–623 (2016).
8. Strawbridge, R. J. et al. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Transl. Psychiatry* **8**, 1–11 (2018).
9. Bulik-Sullivan, B. K. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
10. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
11. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
12. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
13. Byrnes, J. P., Miller, D. C. & Schafer, W. D. Gender differences in risk taking: a meta-analysis. *Psychol. Bull.* **125**, 367–383 (1999).
14. Croson, R. & Gneezy, U. Gender differences in preferences. *J. Econ. Lit.* **47**, 448–474 (2009).
15. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
16. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
17. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–139 (2008).
18. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).
19. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
20. Einav, B. L., Finkelstein, A., Pascu, I. & Cullen, M. R. How general are risk preferences? Choices under uncertainty in different domains. *Am. Econ. Rev.* **102**, 2606–2638 (2016).
21. Frey, R., Pedroni, A., Mata, R., Rieskamp, J. & Hertwig, R. Risk preference shares the psychometric structure of major psychological traits. *Sci. Adv.* **3**, e1701381 (2017).
22. Weber, E. U., Blais, A. E. & Betz, N. E. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *J. Behav. Decis. Mak.* **15**, 263–290 (2002).
23. Hanoch, Y., Johnson, J. G. & Wilke, A. Domain specificity in experimental measures and participant recruitment: an application to risk-taking behavior. *Psychol. Sci.* **17**, 300–304 (2006).
24. Stringer, S. et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32,330 subjects from the International Cannabis Consortium. *Transl. Psychiatry* **6**, e769 (2016).
25. Becker, A., Deckers, T., Dohmen, T., Falk, A. & Kosse, F. The relationship between economic preferences and psychological personality measures. *Annu. Rev. Econ.* **4**, 453–478 (2012).
26. Krueger, R. F. et al. Etiologic connections among substance dependence, antisocial behavior and personality: modeling the externalizing spectrum. *J. Abnorm. Psychol.* **111**, 411–424 (2002).
27. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
28. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* **3**, e3395 (2008).
29. de Vlaming, R. et al. Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLoS Genet.* **13**, e1006495 (2017).
30. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
31. Fehrmann, R. S. N. et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
32. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
33. Petroff, O. A. C. GABA and glutamate in the human brain. *Neurosci.* **8**, 562–573 (2002).
34. Lee, J. et al. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat. Genet.* **50**, 1112–1121 (2018).
35. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
36. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
37. Haber, S. N. & Knutson, B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
38. Tobler, P. N. & Weber, E. U. in *Neuroeconomics* 149–172 (Elsevier, Amsterdam, 2014).
39. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
40. Sahm, C. R. How much does risk tolerance change? *Q. J. Finance* **2**, 1250020 (2012).
41. Malmendier, U. & Nagel, S. Depression babies: do macroeconomic experiences affect risk taking? *Q. J. Econ.* **126**, 373–416 (2011).
42. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
43. Furberg, H. et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).

## Acknowledgements

This research was carried out under the auspices of the Social Science Genetic Association Consortium. The research was also conducted using the UK Biobank Resource under application number 11425. The study was supported by funding from the Ragnar Söderberg Foundation (E9/11 and E42/15); the Swedish Research Council (421-2013-1061); the Jan Wallander and Tom Hedelius Foundation; an ERC Consolidator Grant to Philipp Koellinger (647648 EdGe); the Pershing Square Fund of the Foundations of Human Behavior; the Open Philanthropy Project; the National Institute on Aging, National Institutes of Health through grants P01-AG005842, P01-AG005842-20S2, P30-AG012810, and T32-AG000186-23 to the National Bureau of Economic Research and R01-AG042568-02 to the University of Southern California; the government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-152); and the Social Sciences and Humanities Research Council of Canada. We thank the International Cannabis Consortium, the eQTLgen Consortium, and the Psychiatric Genomics Consortium for sharing summary statistics from the GWAS of lifetime cannabis use, eQTL summary statistics, and summary statistics from the GWAS of ADHD, respectively. A full list of acknowledgments is provided in the Supplementary Note.

## Author contributions

J.P.B., P.D.K., and D.J.B. designed and oversaw the study. J.P.B. closely supervised the analyses and led the writing of the manuscript. R.K.L. was the lead analyst, responsible for quality control, the meta-analyses, summarizing the overlap across the results of the various GWAS, and the SNP heritability analyses. E.K. conducted the population stratification, replication, and proxy-phenotype analyses. R.W. led the genetic correlation analyses, and M.A.F. contributed to those analyses. P.B. led the polygenic score prediction analyses, and R.K.L., E.K., R.W., A.A., R.d.V., M.A.F., and M.G.N. contributed to those analyses. P.B. and C.L.Z. conducted the MTAG analyses. The bioinformatics analyses

were led by S.F.W.M. Analysts who assisted S.F.W.M. include J.G. and M.T. (SMR), A.R.H. (MAGMA), G.A.M. (Gene Network), and P.N.T. (DEPICT). M.L. conducted the review of the literature attempting to link risk tolerance to biological pathways. R.K.L. prepared the majority of figures; E.K. and S.P.T. also prepared some figures. A.O., C.A.R., and S.P.T. helped with several additional analyses. D.C., J.G., and J.J.L. provided helpful advice and feedback on various aspects of the study design. All authors contributed to and critically reviewed the manuscript. P.B., R.K.L., E.K., S.F.W.M., and R.W. made especially major contributions to the writing and editing; these authors contributed equally, and lead authors 2–5 are listed aphabetically in the author list.

## Competing interests

A. Auton, P.F., D.A.H., and A.K. are employees of 23andMe. R.C.K., in the past three years, received support for his epidemiological studies from Sanofi Aventis; was a consultant for Johnson & Johnson Wellness and Prevention, Sage Pharmaceuticals, Shire, and Takeda; and served on an advisory board for the Johnson & Johnson Services Inc. Lake Nona Life Project. R.C.K. is a co-owner of DataStat Inc., a market research firm that carries out healthcare research. J.M. is a principal in BEAM Diagnostics Inc. The authors declare no other competing interests.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41588-018-0309-3.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to R.K. or J.P.B.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Richard Karlsson Linnér[1,2,3,90]*, Pietro Biroli[4,90], Edward Kong[5,90], S. Fleur W. Meddens[1,2,3,90], Robbee Wedow[6,7,8,9,90], Mark Alan Fontana[10,11], Maël Lebreton[12,13], Stephen P. Tino[14], Abdel Abdellaoui[15], Anke R. Hammerschlag[1], Michel G. Nivard[15], Aysu Okbay[1,3], Cornelius A. Rietveld[2,16,17], Pascal N. Timshel[18,19], Maciej Trzaskowski[20], Ronald de Vlaming[1,2,3], Christian L. Zünd[4], Yanchun Bao[21], Laura Buzdugan[22,23], Ann H. Caplin[24], Chia-Yen Chen[6,8,25], Peter Eibich[26,27,28], Pierre Fontanillas[29], Juan R. Gonzalez[30,31,32], Peter K. Joshi[33], Ville Karhunen[34], Aaron Kleinman[29], Remy Z. Levin[35], Christina M. Lill[36], Gerardus A. Meddens[37], Gerard Muntané[38,39,40], Sandra Sanchez-Roige[41], Frank J. van Rooij[17], Erdogan Taskesen[1], Yang Wu[20], Futao Zhang[20], 23and Me Research Team[42], eQTLgen Consortium[42], International Cannabis Consortium[43], Social Science Genetic Association Consortium[42], Adam Auton[29], Jason D. Boardman[44,45,46], David W. Clark[33], Andrew Conlin[47], Conor C. Dolan[15], Urs Fischbacher[48,49], Patrick J. F. Groenen[2,50], Kathleen Mullan Harris[51,52], Gregor Hasler[53], Albert Hofman[7,17], Mohammad A. Ikram[17], Sonia Jain[54], Robert Karlsson[55], Ronald C. Kessler[56], Maarten Kooyman[57], James MacKillop[58,59], Minna Männikkö[34], Carlos Morcillo-Suarez[38], Matthew B. McQueen[60], Klaus M. Schmidt[61], Melissa C. Smart[21], Matthias Sutter[62,63,64], A. Roy Thurik[2,65], André G. Uitterlinden[66], Jon White[67], Harriet de Wit[68], Jian Yang[20,69], Lars Bertram[36,70,71], Dorret I. Boomsma[15], Tõnu Esko[72], Ernst Fehr[23], David A. Hinds[29], Magnus Johannesson[73], Meena Kumari[21], David Laibson[5], Patrik K. E. Magnusson[55], Michelle N. Meyer[74], Arcadi Navarro[38,75,76], Abraham A. Palmer[41,77], Tune H. Pers[18,19], Danielle Posthuma[1,78], Daniel Schunk[79], Murray B. Stein[41,54], Rauli Svento[47], Henning Tiemeier[17], Paul R. H. J. Timmers[33], Patrick Turley[6,8,80], Robert J. Ursano[81], Gert G. Wagner[27,82], James F. Wilson[33,83], Jacob Gratten[20,84], James J. Lee[85], David Cesarini[86], Daniel J. Benjamin[80,87,88,91], Philipp D. Koellinger[1,3,89,91] and Jonathan P. Beauchamp[14,91]*

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, the Netherlands. [2]Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [3]Department of Economics, VU University Amsterdam, Amsterdam, the Netherlands. [4]Department of Economics, University of Zurich, Zurich, Switzerland. [5]Department of Economics, Harvard University, Cambridge, MA, USA. [6]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [8]Analytic Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [9]Department of Sociology, Harvard University, Cambridge, MA, USA. [10]Center for the Advancement of Value in Musculoskeletal Care, Hospital for Special Surgery, New York, NY, USA. [11]Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University, New York, NY, USA. [12]Amsterdam School of Economics, University of Amsterdam, Amsterdam, the Netherlands. [13]Amsterdam Brain and Cognition, University of Amsterdam, Amsterdam, the Netherlands. [14]Department of Economics, University of Toronto, Toronto, Ontario, Canada. [15]Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [16]Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [17]Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands. [18]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark. [19]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [20]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia. [21]Institute for Social and Economic Research, University of Essex, Colchester, UK. [22]Seminar for Statistics, Department of Mathematics, ETH Zurich, Zurich, Switzerland. [23]Department of Economics, University of Zurich, Zurich, Switzerland. [24]Stuyvesant High School, New York, NY, USA. [25]Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [26]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [27]Socio-Economic Panel Study, DIW Berlin, Berlin, Germany. [28]Max Planck Institute for Demographic Research, Rostock, Germany. [29]Research, 23andMe, Inc., Mountain View, CA, USA. [30]Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain. [31]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [32]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [33]Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland. [34]Center for Life Course Health Research, University of Oulu, Oulu, Finland. [35]Department of Economics, University of California San Diego, La Jolla, CA, USA. [36]Genetic and Molecular Epidemiology Group, Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics & Cardiogenetics, University of Lübeck, Lübeck, Germany. [37]Team Loyalty BV, Hoofddorp, the Netherlands. [38]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain. [39]Research Department, Hospital Universitari Institut Pere Mata, Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain. [40]Centro de Investigación Biomédica en Red en Salud Mental (CIBERSAM), Reus, Spain. [41]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [42]A list of members and affiliations appears at the end of the paper. [43]A list of members and affiliations appears in the Supplementary Note. [44]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA. [45]Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, USA. [46]Department of Sociology, University of Colorado Boulder, Boulder, CO, USA. [47]Department of Economics and Finance, Oulu Business School, University of Oulu, Oulu, Finland. [48]Department of Economics, University of Konstanz, Konstanz, Germany. [49]Thurgau Institute of Economics, Kreuzlingen, Switzerland. [50]Department of Econometrics, Erasmus University Rotterdam, Rotterdam, Rotterdam, the Netherlands. [51]Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [52]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [53]Unit of Psychiatry Research, University of Fribourg, Fribourg, Switzerland. [54]Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA. [55]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [56]Department of Health Care Policy, Harvard Medical School, Boston, MA, USA. [57]SURFsara, Amsterdam, the Netherlands. [58]Peter Boris Centre for Addictions Research, McMaster University/St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada. [59]Homewood Research Institute, Guelph, Ontario, Canada. [60]Department of Integrative Physiology, University of Colorado Boulder, Boulder, CO, USA. [61]Department of Economics, University of Munich, Munich, Germany. [62]Department of Economics, University of Cologne, Cologne, Germany. [63]Experimental Economics Group, Max Planck Institute for Research into Collective Goods, Bonn, Germany. [64]Department of Public Finance, University of Innsbruck, Innsbruck, Austria. [65]Montpellier Business School, Montpellier, France. [66]Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, the Netherlands. [67]UCL Genetics Institute, Department of Genetics, Evolution and Environment, University College London, London, UK. [68]Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. [69]Queensland Brain Institute, The University of Queensland, Brisbane, Australia. [70]Dept of Psychology, University of Olso, Oslo, Norway. [71]School of Public Health, Imperial College London, London, UK. [72]Estonian Genome Center, University of Tartu, Tartu, Estonia. [73]Department of Economics, Stockholm School of Economics, Stockholm, Sweden. [74]Center for Translational Bioethics and Health Care Policy, Geisinger Health System, Danville, PA, USA. [75]Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain. [76]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain. [77]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [78]Department of Clinical Genetics, VU Medical Centre, Amsterdam, the Netherlands. [79]Department of Economics, Johannes Gutenberg University, Mainz, Germany. [80]Behavioral and Health Genomics Center, Center for Economic and Social Research, University of Southern California, Los Angeles, CA, USA. [81]Department of Psychiatry, Center for the Study of Traumatic Stress, Uniformed Services University of the Health Sciences, Bethesda, MD, USA. [82]Max Planck Institute for Human Development, Berlin, Germany. [83]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, Scotland. [84]Mater Medical Research Institute, Translational Research Institute, Brisbane, Australia. [85]Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN, USA. [86]Department of Economics, New York University, New York, NY, USA. [87]Department of Economics, University of Southern California, Los Angeles, CA, USA. [88]National Bureau of Economic Research, Cambridge, MA, USA. [89]German Institute for Economic Research, DIW Berlin, Berlin, Germany. [90]These authors contributed equally: Richard Karlsson Linnér, Pietro Biroli, Edward Kong, S. Fleur W. Meddens, Robbee Wedow. [91]These authors jointly supervised this work: Daniel J. Benjamin, Philipp D. Koellinger, Jonathan P. Beauchamp. *e-mail: r.karlssonlinner@vu.nl; jonathan.pierre.beauchamp@gmail.com

**23and Me Research Team**

Michelle Agee[29], Babak Alipanahi[29], Robert K. Bell[29], Katarzyna Bryc[29], Sarah L. Elson[29], Nicholas A. Furlotte[29], Karen E. Huber[29], Nadia K. Litterman[29], Jennifer C. McCreight[29], Matthew H. McIntyre[29], Joanna L. Mountain[29], Carrie A. M. Northover[29], Steven J. Pitts[29], J. Fah Sathirapongsasuti[29], Olga V. Sazonova[29], Janie F. Shelton[29], Suyash Shringarpure[29], Chao Tian[29], Joyce Y. Tung[29], Vladimir Vacic[29] and Catherine H. Wilson[29]

**eQTLgen Consortium**

Mawussé Agbessi[92], Habibul Ahsan[93], Isabel Alves[92], Anand Andiappan[94], Philip Awadalla[92], Alexis Battle[95], Frank Beutner[96], Marc Jan Bonder[97], Dorret I. Boomsma[15], Mark Christiansen[98], Annique Claringbould[97], Patrick Deelen[97], Tõnu Esko[72], Marie-Julie Favé[92], Lude Franke[97], Timothy Frayling[99], Sina A. Gharib[100,101], Gregory Gibson[102], Bastiaan Heijmans[103], Gibran Hemani[104], Rick Jansen[105], Mika Kähönen[106], Anette Kalnapenkis[72], Silva Kasela[72], Johannes Kettunen[107], Yungil Kim[95,108], Holger Kirsten[109], Peter Kovacs[110], Knut Krohn[111], Jaanika Kronberg-Guzman[72], Viktorija Kukushkina[72], Zoltan Kutalik[112], Bernett Lee[94], Terho Lehtimäki[113], Markus Loeffler[109], Urko M. Marigorta[102], Andres Metspalu[72], Lili Milani[72], Grant W. Montgomery[20], Martina Müller-Nurasyid[114], Matthias Nauck[115], Michel G. Nivard[105], Brenda Penninx[105], Markus Perola[116], Natalia Pervjakova[72], Brandon Pierce[93], Joseph Powell[117], Holger Prokisch[118], Bruce M. Psaty[98], Olli Raitakari[119], Susan Ring[104,120], Samuli Ripatti[107], Olaf Rotzchke[94], Sina Rüeger[112], Ashis Saha[95], Markus Scholz[109], Katharina Schramm[114], Ilkka Seppälä[113], Michael Stumvoll[121], Patrick Sullivan[55], Peter-Bram t Hoen[122], Alexander Teumer[123], Joachim Thiery[124], Lin Tong[93], Anke Tönjes[121], Jenny van Dongen[105], Joyce van Meurs[66], Joost Verlouw[66], Peter M. Visscher[20], Uwe Völker[125], Urmo Võsa[97], Harm-Jan Westra[97], Hanieh Yaghootkar[99], Jian Yang[20,69], Biao Zeng[102] and Futao Zhang[20]

**Social Science Genetic Association Consortium**

Aysu Okbay[1,3], Jonathan P. Beauchamp[14], Mark Alan Fontana[10,11], James J. Lee[85], Tune H. Pers[18,19], Cornelius A. Rietveld[2,16,17], Patrick Turley[6,8,80], Guo-Bo Chen[69], Valur Emilsson[126,127], S. Fleur W. Meddens[1,2,3], Sven Oskarsson[128], Joseph K. Pickrell[129], Kevin Thom[86], Pascal Timshel[18,19], Ronald de Vlaming[1,2,3], Abdel Abdellaoui[15], Tarunveer S. Ahluwalia[18,130,131], Jonas Bacelis[132], Clemens Baumbach[133,134], Gyda Bjornsdottir[135], Johannes H. Brandsma[136], Maria Pina Concas[137], Jaime Derringer[138], Nicholas A. Furlotte[29], Tessel E. Galesloot[122], Giorgia Girotto[139], Richa Gupta[140], Leanne M. Hall[141,142], Sarah E. Harris[143,144], Edith Hofer[145,146], Momoko Horikoshi[147,148], Jennifer E. Huffman[83], Kadri Kaasik[149], Ioanna P. Kalafati[150], Robert Karlsson[55], Augustine Kong[135], Jari Lahti[149,151], Sven J. van der Lee[17], Christiaan de Leeuw[1,152], Penelope A. Lind[153], Karl-Oskar Lindgren[128], Tian Liu[82], Massimo Mangino[154,155], Jonathan Marten[83], Evelin Mihailov[72], Michael B. Miller[85], Peter J. van der Most[156], Christopher Oldmeadow[157,158], Antony Payton[159,160], Natalia Pervjakova[72], Wouter J. Peyrot[161,162], Yong Qian[163], Olli Raitakari[119], Rico Rueedi[164,165], Erika Salvi[166], Börge Schmidt[167], Katharina E. Schraut[33], Jianxin Shi[168], Albert V. Smith[126,169], Raymond A. Poot[136], Beate St Pourcain[104,170], Alexander Teumer[123], Gudmar Thorleifsson[135], Niek Verweij[171], Dragana Vuckovic[139], Juergen Wellmann[172], Harm-Jan Westra[97], Jingyun Yang[173,174], Wei Zhao[175], Zhihong Zhu[69],

Behrooz Z. Alizadeh[156,176], Najaf Amin[17], Andrew Bakshi[69], Sebastian E. Baumeister[123,177], Ginevra Biino[178], Klaus Bønnelykke[130], Patricia A. Boyle[173,179], Harry Campbell[33], Francesco P. Cappuccio[180], Gail Davies[143,181], Jan-Emmanuel De Neve[182], Panos Deloukas[183,184], Ilja Demuth[185,186], Jun Ding[163], Peter Eibich[26,27,28], Lewin Eisele[167], Niina Eklund[116], David M. Evans[104,187], Jessica D. Faul[188], Mary F. Feitosa[189], Andreas J. Forstner[190,191], Ilaria Gandin[139], Bjarni Gunnarsson[137], Bjarni V. Halldórsson[135,192], Tamara B. Harris[193], Andrew C. Heath[194], Lynne J. Hocking[195], Elizabeth G. Holliday[157,158], Georg Homuth[125], Michael A. Horan[196], Jouke-Jan Hottenga[15], Philip L. de Jager[197,198,199], Peter K. Joshi[33], Astanand Jugessur[200], Marika A. Kaakinen[201], Mika Kähönen[106], Stavroula Kanoni[183], Liisa Keltigangas-Järvinen[149], Lambertus A. L. M. Kiemeney[122], Ivana Kolcic[202], Seppo Koskinen[116], Aldi T. Kraja[189], Martin Kroh[27], Zoltan Kutalik[112], Antti Latvala[140], Lenore J. Launer[203], Maël P. Lebreton[12,13], Douglas F. Levinson[204], Paul Lichtenstein[55], Peter Lichtner[118], David C. M. Liewald[143,181], Anu Loukola[140], Pamela A. Madden[194], Reedik Mägi[72], Tomi Mäki-Opas[116], Riccardo E. Marioni[69,143,205], Pedro Marques-Vidal[206], Gerardus A. Meddens[37], George McMahon[104], Christa Meisinger[134], Thomas Meitinger[118], Yusplitri Milaneschi[161,162], Lili Milani[72], Grant W. Montgomery[20], Ronny Myhre[200], Christopher P. Nelson[141,142], Dale R. Nyholt[207,208], William E. R. Ollier[159], Aarno Palotie[6,25,197,209,210,211], Lavinia Paternoster[104], Nancy L. Pedersen[55], Katja E. Petrovic[145], David J. Porteous[144], Katri Räikkönen[149,151], Susan M. Ring[104], Antonietta Robino[212], Olga Rostapshova[5,213], Igor Rudan[33], Aldo Rustichini[214], Veikko Salomaa[116], Alan R. Sanders[68,215], Antti-Pekka Sarin[210,216], Helena Schmidt[145,217], Rodney J. Scott[158,218], Blair H. Smith[219], Jennifer A. Smith[175], Jan A. Staessen[220,221], Elisabeth Steinhagen-Thiessen[185], Konstantin Strauch[114,222], Antonio Terracciano[223], Martin D. Tobin[224], Sheila Ulivi[212], Simona Vaccargiu[137], Lydia Quaye[154], Frank J. van Rooij[17], Cristina Venturini[154,155], Anna A. E. Vinkhuyzen[69], Uwe Völker[125], Henry Völzke[123], Judith M. Vonk[156], Diego Vozzi[213], Johannes Waage[130,131], Erin B. Ware[175,225], Gonneke Willemsen[15], John R. Attia[157,158], David A. Bennett[173,174], Klaus Berger[171], Lars Bertram[36,70,71], Hans Bisgaard[130], Dorret I. Boomsma[15], Ingrid B Borecki[189], Ute Bültmann[226], Christopher F. Chabris[227], Francesco Cucca[228], Daniele Cusi[166,229], Ian J. Deary[143,181], George V. Dedoussis[150], Cornelia M. van Duijn[17], Johan G. Eriksson[151,230], Barbara Franke[231], Lude Franke[97], Paolo Gasparini[139,212,232], Pablo V. Gejman[68,215], Christian Gieger[133], Hans-Jörgen Grabe[233,234], Jacob Gratten[20,84], Patrick J. F. Groenen[2,50], Vilmundur Gudnason[126,169], Pim van der Harst[97,171,235], Caroline Hayward[83,236], David A. Hinds[29], Wolfgang Hoffmann[123], Elina Hyppönen[237,238,239], William G. Iacono[85], Bo Jacobsson[132,200], Marjo-Riitta Järvelin[240,241,242,243], Karl-Heinz Jöckel[167], Jaakko Kaprio[116,140,210], Sharon L. R. Kardia[175], Terho Lehtimäki[113], Steven F. Lehrer[244,245], Patrik K. E. Magnusson[55], Nicholas G. Martin[246], Matt McGue[85], Andres Metspalu[72], Neil Pendleton[247,248], Brenda Penninx[105], Markus Perola[116], Nicola Pirastu[139], Mario Pirastu[137], Ozren Polasek[33,249], Danielle Posthuma[1,78], Christine Power[239], Michael A. Province[189], Nilesh J. Samani[141,142], David Schlessinger[163], Reinhold Schmidt[145], Thorkild I. A. Sørensen[18,104,250], Tim D. Spector[154], Kari Stefansson[135,169], Unnur Thorsteinsdottir[135,169], A. Roy Thurik[2,65], Nicholas J. Timpson[104], Henning Tiemeier[17], Joyce Y. Tung[29], André G. Uitterlinden[66], Veronique Vitart[83], Peter Vollenweider[206], David R. Weir[188], James F. Wilson[33,83], Alan F. Wright[83], Dalton C. Conley[251,252], Robert F. Krueger[85], George Davey Smith[104], Albert Hofman[7,17], David I. Laibson[5], Sarah E. Medland[153], Michelle N. Meyer[74], Jian Yang[20,69], Magnus Johannesson[73], Peter M. Visscher[20], Tõnu Esko[72], Philipp D. Koellinger[1,3,89], David Cesarini[86] and Daniel J. Benjamin[80,87,88]

[92]Computational Biology, Ontario Institute for Cancer Research, Toronto, Canada. [93]Department of Public Health Sciences, University of Chicago, Chicago, IL, USA. [94]Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore. [95]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. [96]Heart Center Leipzig, Universität Leipzig, Leipzig, Germany. [97]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [98]Cardiovascular Health Research Unit, Departments of Epidemiology, Medicine, and Health Services, University of Washington, Seattle, WA, USA. [99]Medical School, University of Exeter, Exeter, UK. [100]Department of Medicine, University of Washington, Seattle, WA, USA. [101]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [102]School of Biological Sciences, Georgia Tech, Atlanta, GA, USA. [103]Leiden University Medical Centre, Leiden, the Netherlands. [104]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [105]Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [106]Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [107]University of Helsinki, Helsinki, Finland. [108]Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [109]Institut für Medizinische Informatik, Statistik und Epidemiologie, Leipzig Research Centre for Civilization Diseases (LIFE), Universität Leipzig, Leipzig, Germany. [110]IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany. [111]Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany. [112]Lausanne University Hospital, Lausanne, Switzerland. [113]Department of Clinical Chemistry, Fimlab Laboratories and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [114]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [115]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [116]National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland. [117]Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, New South Wales, Australia. [118]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [119]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital and University of Turku, Turku, Finland. [120]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [121]Department of Medicine, Universität Leipzig, Leipzig, Germany. [122]Radboud University Medical Center, Radboud University, Nijmegen, the Netherlands. [123]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [124]Institute for Laboratory Medicine, Leipzig Research Center for Civilization Diseases (LIFE), Universität Leipzig, Leipzig, Germany. [125]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [126]Icelandic Heart Association, Kopavogur, Iceland. [127]Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavík, Iceland. [128]Department of Government, Uppsala University, Uppsala, Sweden. [129]New York Genome Center, New York, NY, USA. [130]Copenhagen Prospective Studies on Asthma in Childhood (COPSAC), Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. [131]Steno Diabetes Center, Gentofte, Denmark. [132]Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, Sweden. [133]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [134]Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [135]deCODE Genetics/Amgen Inc., Reykjavik, Iceland. [136]Department of Cell Biology, Erasmus Medical Center, Rotterdam, the Netherlands. [137]Istituto di Ricerca Genetica e Biomedica UOS di Sassari, National Research Council of Italy, Sassari, Italy. [138]Psychology, University of Illinois, Champaign, IL, USA. [139]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy. [140]Department of Public Health, University of Helsinki, Helsinki, Finland. [141]Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. [142]NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [143]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [144]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [145]Department of Neurology, General Hospital and Medical University Graz, Graz, Austria. [146]Institute for Medical Informatics, Statistics and Documentation, General Hospital and Medical University Graz, Graz, Austria. [147]Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, UK. [148]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [149]Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland. [150]Nutrition and Dietetics, Health Science and Education, Harokopio University, Athens, Greece. [151]Folkhälsan Research Centre, Helsingfors, Finland. [152]Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, the Netherlands. [153]Quantitative Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [154]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [155]NIHR Biomedical Research Centre, Guy's and St. Thomas' Foundation Trust, London, UK. [156]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [157]Public Health Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia. [158]Faculty of Health and Medicine, University of Newcastle, Newcastle, New South Wales, Australia. [159]Centre for Integrated Genomic Medical Research, Institute of Population Health, University of Manchester, Manchester, UK. [160]Human Communication and Deafness, School of Psychological Sciences, University of Manchester, Manchester, UK. [161]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands. [162]GGZ inGeest, Amsterdam, the Netherlands. [163]Laboratory of Genetics, National Institute on Aging, Baltimore, MD, USA. [164]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [165]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [166]Department Of Health Sciences, University of Milan, Milano, Italy. [167]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, Germany. [168]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [169]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [170]School of Oral and Dental Sciences, University of Bristol, Bristol, UK. [171]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [172]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [173]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [174]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA. [175]Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA. [176]Department of Gastroenterology and Hepatology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [177]Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany. [178]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. [179]Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL, USA. [180]Warwick Medical School, University of Warwick, Coventry, UK. [181]Department of Psychology, University of Edinburgh, Edinburgh, UK. [182]Saïd Business School, University of Oxford, Oxford, UK. [183]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [184]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [185]The Berlin Aging Study II Research Group on Geriatrics, Charité–Universitätsmedizin Berlin, Berlin, Germany. [186]Institute of Medical and Human Genetics, Charité–Universitätsmedizin Berlin, Berlin, Germany. [187]University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, Queensland, Australia. [188]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [189]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [190]Institute of Human Genetics, University of Bonn, Bonn, Germany. [191]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [192]Institute of Biomedical and Neural Engineering, School of Science and Engineering, Reykjavik University, Reykjavik, Iceland. [193]Laboratory of Epidemiology & Population Science, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [194]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [195]Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK. [196]Manchester Medical School, University of Manchester, Manchester, UK. [197]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [198]Program in Translational NeuroPsychiatric Genomics, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. [199]Harvard Medical School, Boston, Massachusetts, USA. [200]Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway. [201]Department of Genomics of

Common Disease, Imperial College London, London, UK. [202]Public Health, Medical School, University of Split, Split, Croatia. [203]Neuroepidemiology Section, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [204]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [205]Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [206]Department of Internal Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [207]Molecular Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [208]Institute of Health and Biomedical Innovation, Queensland Institute of Technology, Brisbane, Queensland, Australia. [209]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [210]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [211]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [212]Medical Genetics, Institute for Maternal and Child Health IRCCS 'Burlo Garofolo', Trieste, Italy. [213]Social Impact, Arlington, VA, USA. [214]Department of Economics, University of Minnesota Twin Cities, Minneapolis, MN, USA. [215]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL, USA. [216]Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland. [217]Research Unit for Genetic Epidemiology, Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, General Hospital and Medical University, Graz, Austria. [218]Information Based Medicine Stream, Hunter Medical Research Institute, New Lambton, New South Wales, Australia. [219]Medical Research Institute, University of Dundee, Dundee, UK. [220]Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Science, University of Leuven, Leuven, Belgium. [221]R&D VitaK Group, Maastricht University, Maastricht, the Netherlands. [222]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians-Universität, Munich, Germany. [223]Department of Geriatrics, Florida State University College of Medicine, Tallahassee, FL, USA. [224]Department of Health Sciences and Genetics, University of Leicester, Leicester, UK. [225]Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [226]Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. [227]Department of Psychology, Union College, Schenectady, NY, USA. [228]Istituto di Ricerca Genetica e Biomedica (IRGB), Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, Cagliari, Italy. [229]Institute of Biomedical Technologies, Italian National Research Council, Milan, Italy. [230]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [231]Departments of Human Genetics and Psychiatry, Donders Centre for Neuroscience, Nijmegen, the Netherlands. [232]Experimental Genetics Division, Sidra, Doha, Qatar. [233]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [234]Department of Psychiatry and Psychotherapy, HELIOS-Hospital Stralsund, Stralsund, Germany. [235]Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, the Netherlands. [236]Generation Scotland, Centre for Genomics and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [237]Centre for Population Health Research, School of Health Sciences and Sansom Institute, University of South Australia, Adelaide, South Australia, Australia. [238]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [239]Population, Policy and Practice, UCL Institute of Child Health, London, UK. [240]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [241]Center for Life Course Epidemiology, Faculty of Medicine, University of Oulu, Oulu, Finland. [242]Unit of Primary Care, Oulu University Hospital, Oulu, Finland. [243]Biocenter Oulu, University of Oulu, Oulu, Finland. [244]Economics, NYU Shanghai, Pudong, China. [245]Policy Studies, Queen's University, Kingston, Ontario, Canada. [246]Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [247]Centre for Clinical and Cognitive Neuroscience, Institute of Brain Behaviour and Mental Health, Salford Royal Hospital, Manchester, UK. [248]Manchester Institute for Collaborative Research in Ageing, University of Manchester, Manchester, UK. [249]Faculty of Medicine, University of Split, Split, Croatia. [250]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospitals, Frederiksberg, Denmark. [251]Department of Sociology, New York University, New York, NY, USA. [252]School of Medicine, New York University, New York, NY, USA.

## Methods

This article is accompanied by a Supplementary Note with further details. Further information on experimental design is also available in the Life Sciences Reporting Summary linked to this article.

**Phenotype definitions, GWAS, quality control, and meta-analysis.** For our discovery GWAS of general risk tolerance ($n=939{,}908$), we performed a sample-size-weighted meta-analysis of results from the UKB ($n=431{,}126$) and a sample of research participants from 23andMe ($n=508{,}782$). For our replication GWAS of general risk tolerance ($n=35{,}445$), we performed a sample-size-weighted meta-analysis of results from ten smaller cohorts from seven studies: Army STARRS, BASE-II, NFBC 1966, RSIII, STR, UKHLS, and VIKING. The exact measures for the general risk tolerance phenotype vary across cohorts in wording and number of response categories, but all measures are similar and ask about one's tendency, preparedness, or willingness to take risks in general (Supplementary Table 4).

For our GWAS of adventurousness, we analyzed data from a sample of research participants from 23andMe ($n=557{,}923$). We analyzed responses to the question 'If forced to choose, would you consider yourself to be more cautious or more adventurous?', with possible responses ranging from '(1) very cautious' to '(5) very adventurous'. For our GWAS of three of the four risky behaviors—automobile speeding propensity ($n=404{,}291$), drinks per week ($n=414{,}343$), and number of sexual partners ($n=370{,}711$)—and for the first PC of the four risky behaviors ($n=315{,}894$), we analyzed UKB data. For the remaining risky behavior, ever smoker ($n=518{,}633$), we meta-analyzed GWAS results from the UKB and from the TAG Consortium[43]. Our automobile speeding propensity phenotype is based on responses to the question 'How often do you drive faster than the speed limit on the motorway?', with possible responses ranging from '(1) never/rarely' to '(4) most of the time'. We dropped individuals who answered '(5) do not drive on the motorway', and then we normalized the categorical variable for males and females separately. Our drinks per week phenotype was constructed on the basis of responses to a series of questions about drinking habits and is defined as the number of alcoholic drinks consumed per week. Our ever-smoker phenotype in the UKB is a dummy variable that was equal to one if a respondent reported being a current or previous smoker and zero if the respondent reported never smoking or only smoking once or twice; our ever smoker phenotype from the TAG Consortium is the consortium's 'smoking initiation' phenotype (which TAG also refers to as 'ever versus never regular smoker')[43]. Our number of sexual partners phenotype is based on responses to the question 'About how many sexual partners have you had in your lifetime?'; respondents who reported more than 99 lifetime sexual partners were asked to confirm their responses. We assigned a value of zero to participants who reported having never had sex, and we again normalized this measure separately for males and females. Our first PC phenotype is the first PC obtained from a PC analysis in the UKB of the four risky behaviors (Supplementary Table 23). All seven phenotypes were coded such that higher phenotype values were associated with higher risk tolerance or risk taking. Table 1 lists, for each GWAS, the datasets we analyzed and the GWAS sample size. The Supplementary Note and Supplementary Tables 4 and 5 provide additional details on the cohorts and phenotype definitions.

All GWAS were performed at the cohort level in European-ancestry subjects according to a prespecified and publicly archived analysis plan (see URLs). All GWAS included controls for the top ten (or more) PCs of the genetic relatedness matrix and for sex and birth year. Genotyping was performed by using a range of commercially available genotyping arrays. We applied extensive quality-control procedures to the cohort-level summary statistics, including but not limited to the EasyQC protocol developed by the GIANT consortium[44]. We used Haplotype Reference Consortium v.1.1 data to construct our main reference panel, which we used for quality control of the GWAS summary statistics and to determine the independence of significant loci. For the 23andMe and UKB cohorts, only SNPs with minor allele frequency (MAF) greater than 0.001 were analyzed. All meta-analyses were restricted to SNPs with a sample size greater than half of the maximum sample size across all the SNPs in the GWAS. In total, 9,284,738 SNPs were analyzed in the discovery GWAS of general risk tolerance; 9,339,358 SNPs were analyzed in the GWAS of adventurousness; and ~11,515,000 SNPS were analyzed in the GWAS of the four risky behaviors and their first PC. To adjust standard errors for the possible effects of population stratification, we inflated them by the square root of the estimated intercept from an LD score regression (for the replication GWAS of general risk tolerance, which meta-analyzed different cohorts, we inflated them at the meta-analysis level). Additional details are provided in the Supplementary Note and Supplementary Tables 2 and 24–26.

To identify approximately independent lead SNPs, we applied a clumping algorithm to the GWAS results. Our clumping algorithm begins by selecting the SNP with the lowest $P$ value as the lead SNP in the first clump and includes in the first clump all SNPs that have $r^2$ greater than 0.1 with the lead SNP and that have a GWAS $P$ value less than $1\times10^{-4}$. Next, the SNP with the second-lowest $P$ value outside the first clump becomes the lead SNP of the second clump, and the second clump is created analogously but using only the SNPs outside of the first clump. This process continues until every genome-wide-significant SNP (that is, every SNP with a GWAS $P$ value less than $5\times10^{-8}$) is either designated as a lead SNP or is clumped to another lead SNP. We also defined nonoverlapping, continuous genomic loci around the lead SNPs by using the locus definition of Ripke et al.[45], and we performed COJO analyses[12]. Ripke et al. defined a locus as "the physical region containing all SNPs correlated at $r^2>0.6$ with [one of the lead] SNPs," and merged associated loci within 250 kb of each other. To define the set of distinct loci containing all the loci corresponding to the locus associations from across the seven GWAS, we pooled the loci corresponding to the locus associations and merged loci within 250 kb of each other. For the COJO analyses, for each of the seven main GWAS we restricted the analysis to the set of SNPs that (i) passed all GWAS quality-control filters and (ii) were located within the loci of the phenotype (which included all of the lead SNPs).

Supplementary Tables 3, 6, 7, and 27 report the lead SNPs, the loci, the results of the COJO analyses, and the results of a lookup of the lead SNPs in the NHGRI-EBI GWAS Catalog database[19] for our seven main GWAS; Supplementary Fig. 12 shows the GWAS estimates of general-risk-tolerance lead SNPs in the 23andMe and UKB cohorts and in the replication GWAS, and the Supplementary Data show LocusZoom plots for all the loci identified in the seven GWAS.

**Testing for population stratification.** To assess the extent to which population stratification might have biased our GWAS estimates, we conducted three tests. First, we estimated LD score intercepts by using the summary statistics of the discovery and replication GWAS of general risk tolerance and of the GWAS of our four main risky behaviors and their first PC. Second, following Okbay et al.[10], we conducted sign tests that compared the signs of the estimates from our discovery GWAS of general risk tolerance (but excluded all full siblings from the UKB cohort) to the signs of the estimates from within-family GWAS of general risk tolerance. If our discovery GWAS estimates were entirely driven by stratification, then the signs of the within-family estimates—which are immune to stratification—should be independent of the signs of the discovery GWAS estimates, in which case we would expect a sign concordance of roughly 50%. A higher degree of sign concordance would suggest that at least some of the signal from the GWAS comes from true genetic effects. Across four sign tests, we strongly rejected the null hypothesis of 50% sign concordance for all of the sign tests ($P<5\times10^{-10}$ in all four tests), thus implying that at least some of the signal from the GWAS came from true genetic effects. Our third test of population stratification, the within-family regression test, compared both the signs and magnitudes of the discovery and within-family GWAS of general risk tolerance. The Supplementary Note, Supplementary Tables 28 and 29, and Supplementary Fig. 13 provide further details on the three tests and report their results. All three tests implied no more than low levels of population stratification.

**Replication of the general-risk-tolerance lead SNPs and maxFDR calculation.** To assess the credibility of the lead SNPs from our discovery GWAS of general risk tolerance, we compared those results with the estimates from our replication GWAS of general risk tolerance. (We did not attempt replication of the results of our six supplementary GWAS in independent data, because we did not have access to such data for these phenotypes.) We first filtered out SNPs with sample size less than one-half the maximum sample size in the replication GWAS. After we applied this filter, 122 of the 124 lead SNPs were directly available in the replication GWAS summary statistics, and one of the two remaining lead SNPs was well proxied by a SNP in high LD ($r^2>0.8$) with it. For the resulting 123 SNPs, we conducted a (one-sided) binomial sign test to assess whether the directions (that is, the signs) of the effects of the lead SNPs were more concordant across the discovery and the replication GWAS than expected by chance. We also conducted a (one-sided) binomial test to assess whether a larger fraction of the lead SNPs was significant at the 5% level in one-sided tests in the replication GWAS than expected by chance. We then followed the procedure outlined by Okbay et al.[46] and conducted a Bayesian analysis to obtain estimates of the posterior distributions of the 123 SNPs' true effect sizes (the $\beta_j$ values), given their GWAS estimates. We used the SNPs' estimated posterior distributions to estimate their expected replication record in the two binomial tests and compared their actual and expected replication records.

To calculate the maxFDR, an upper bound on the FDR for a GWAS, we used the MTAG software[15] and followed the methodology described in section 1.4.3 in the supplementary information of Turley et al.[15]. The maxFDR is defined as the maximum theoretical FDR over a range of possible fractions of null SNPs ($\pi_{\text{null}}$).

The Supplementary Note and Supplementary Fig. 3 provide additional details.

**Estimation of genome-wide SNP heritability.** We used the heritability estimator from summary statistics (HESS)[47] method to estimate the genome-wide SNP heritability of our seven main phenotypes. For the results reported in Table 1, we used the summary statistics from the GWAS listed in the table for all 1000 Genomes phase 3 SNPs with MAF greater than 0.05. We did not apply genomic control (GC) prior to estimating heritability with HESS. The Supplementary Note, Supplementary Table 30, and Supplementary Fig. 14 provide additional details and also report estimates of the SNP heritability of our seven main phenotypes estimated with the GCTA[48], LD score regression, and HESS methods, using only summary statistics from the UKB GWAS for comparability across phenotypes and methods (except for adventurousness, which is not available in the UKB and for which we used the 23andMe summary statistics).

**Genetic correlations.** We used bivariate LD score regression[9] to estimate genetic correlations between general risk tolerance and various phenotypes. We used the scores computed by Finucane et al.[49], which are based on genotypic data from the European-ancestry samples in the 1000 Genomes Project and only HapMap3 SNPs. As is common in the literature, we restricted our analyses to SNPs with MAF >0.01. We used the summary statistics of the meta-analysis combining our discovery and replication GWAS of general risk tolerance to estimate genetic correlations with general risk tolerance, and we used the summary statistics of our GWAS of adventurousness, our four main risky behaviors, and their first PC to estimate genetic correlations with those phenotypes. For most other phenotypes, we used published GWAS results. We obtained the summary statistics from the GWAS of lifetime cannabis use[24] and of ADHD[50] from the International Cannabis Consortium and the Psychiatric Genomics Consortium, respectively. We conducted our own GWAS using the first release of the UKB data for five phenotypes: age first had sexual intercourse ($n = 98{,}956$), teenage conception among females ($n = 40{,}077$), use of sun protection ($n = 111{,}560$), household income ($n = 97{,}059$), and Townsend deprivation index score ($n = 112{,}192$). The sex-specific GWAS of general risk tolerance used to estimate the genetic correlation between males and females were conducted in the full release of UKB data, separately for males and females, by following the same methodology and quality-control protocol as for our other GWAS in the full release of UKB data. The Supplementary Note and Supplementary Tables 9 and 31 provide additional details. In addition, the Supplementary Note, Supplementary Table 32, and Supplementary Fig. 15 report the results of proxy-phenotype analyses in which we examined whether the general-risk-tolerance lead SNPs tended to also be associated with related phenotypes.

**MTAG.** We used MTAG[15] to increase the precision of our estimates of the SNPs' effects on general risk tolerance. We used as inputs the summary statistics of the meta-analysis combining our discovery and replication GWAS of general risk tolerance; the summary statistics of our GWAS of adventurousness, automobile speeding propensity, drinks per week, ever smoker, and number of sexual partners; and the summary statistics of a previously published GWAS on lifetime cannabis use[51]. Because SNPs that have no effect on one phenotype but a sizeable effect on another can bias MTAG results, we excluded from this analysis SNPs in the proximity of several genes implicated in biological processes likely to be specific to only one of the phenotypes. Specifically, we excluded all SNPs located within 1 Mb of the genes *CHRNA5* and *CHRNB3* (nicotinic receptors), *CNR1* and *CNR2* (cannabinoid receptors), and *ADH1B* (alcohol dehydrogenase). We imposed a MAF filter of 0.01 and a sample-size filter that selected, for each GWAS, the SNPs with sample sizes larger than two-thirds of the ninth decile of the GWAS's sample size. MTAG limited the analysis to the 5,869,552 SNPs analyzed in all GWAS (and that satisfied these filters). To identify approximately independent lead SNPs for general risk tolerance, we applied the clumping algorithm described above. The Supplementary Note, Supplementary Table 10, and Supplementary Fig. 7 provide further details.

**Polygenic prediction.** We assessed the predictive power of polygenic scores of general risk tolerance in six different validation cohorts: Add Health, HRS, NTR, STR, UKB-siblings, and Zurich. (The UKB-siblings cohort comprised all individuals with at least one full sibling in the UKB.) We constructed three polygenic scores. Our first two polygenic scores were constructed with the LDpred[27] method, which accounts for the LD between SNPs. The first used the summary statistics from the meta-analysis of the discovery and replication GWAS of general risk tolerance, whereas the second used the MTAG summary statistics. (The LDpred method relies on a Gaussian mixture weight that corresponds to the assumed fraction of SNPs that are causal. For each of our first two polygenic scores, we first generated LDpred scores for each of the following mixture weights: 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 0.0003, and 0.0001 (ref. [52]). The LDpred score results presented in this paper for our first two polygenic scores are for the scores based on a Gaussian mixture weight of 0.3 (our 'preferred score'), which consistently performed well across cohorts and phenotypes.) Our third polygenic score was constructed with the classical method, which simply weights SNPs by their GWAS effect size[53,54], using the summary statistics from the meta-analysis of the discovery and replication GWAS of general risk tolerance.

We used the subset of all the SNPs (that is, we did not impose a *P*-value threshold) in the HapMap consortium phase 3 release[55] with an imputation quality of more than 0.7 to generate all three scores. For every validation cohort that was also included in the discovery or replication GWAS or in the MTAG analysis, we reran the GWAS and MTAG analyses without the validation cohort to generate the summary statistics we used to construct the scores. Owing to data access limitations, the 23andMe cohort could not be included in the meta-analysis whose summary statistics we used to construct the polygenic scores in the NTR, STR, and Zurich cohorts. The second polygenic score (using the MTAG summary statistics) was constructed only for the Add Health, HRS, and UKB-siblings cohorts.

Our measure of a score's predictive power for a predicted phenotype is the incremental $R^2$ (or incremental pseudo-$R^2$) from adding the score to a regression of the phenotype on controls for sex, birth year, birth year squared, and birth year cubed, as well as the interactions between sex and the three birth year variables, and the first ten principal components of the genetic relatedness matrix. We used the bootstrap method with 1,000 iterations to estimate the ninety-fifth-percentile confidence intervals for the incremental $R^2$ estimates. For continuous phenotypes, we estimated ordinary least squares regressions; for binary phenotypes (for example, ever smoker), we estimated probit models; and for censored phenotypes (for example, equity share, which is non-negative), we estimated tobit models. For binary and censored phenotypes, we used McFadden's pseudo-$R^2$ to calculate the incremental pseudo-$R^2$.

The Supplementary Note provides additional details, including a description of how the predicted phenotypes were constructed. Results are presented in Supplementary Figs. 8 and 9 and Supplementary Tables 11–14.

**Biological annotation: testing hypotheses about specific genes and gene sets.** We conducted a comprehensive review of the literature on biological pathways that have been hypothesized to influence risk tolerance. The 132 articles identified by the review are compiled in Supplementary Table 1. The Supplementary Note and Supplementary Tables 15, 16, 33, and 34 provide further details and report the results of the various analyses we conducted to assess whether the pathways and genes that have previously been hypothesized to relate to risk tolerance do indeed show evidence of association with risk tolerance.

**Biological annotation: additional bioinformatics analyses.** We conducted a series of additional bioinformatics analyses using the results of the combined meta-analysis of our discovery and replication GWAS of general risk tolerance. We conducted a gene analysis with MAGMA[30] to test each of 18,224 genes for association with general risk tolerance in a hypothesis-free manner (the 18,224 genes are the set of all genes containing at least one SNP in our combined meta-analysis results). We used our main reference panel to estimate LD. Bonferroni correction was applied to account for multiple testing, counting each gene as an independent test. We then used the Gene Network[31] co-expression database to gain insight into the functions of the significant MAGMA genes.

We also used DEPICT[32] (release 194) to prioritize tissues, gene sets, and genes implicated by our GWAS results. Only SNPs with GWAS *P* values less than $10^{-5}$ were used as input, and DEPICT-defined loci were defined by clumping these SNPs (clumping parameters used for this analysis are shown in the Supplementary Note). Locus boundaries were then defined by using a LD $r^2$ threshold of 0.5, and overlapping loci were merged, yielding 464 autosomal loci comprising 1,060 genes.

To partition the SNP-based heritability of general risk tolerance, we used stratified LD score regression[49], following the procedure described by Finucane et al.[49]. We estimated stratified LD score regressions both for the functional genomic regions of the 'baseline model' and for the tissue-level annotations provided by Finucane et al. To correct for multiple hypothesis testing, we applied a Bonferroni correction for 52 two-sided tests in the baseline model (that is, for 52 annotations) and for ten two-sided tests in the tissue type models (that is, for ten tissue types).

The Supplementary Note, Supplementary Tables 17–22 and 35–39, and Supplementary Figs. 10, 11, and 16 provide further details and report the results of these and other bioinformatics analyses, including a transcriptome-wide analysis with summary-based Mendelian randomization (SMR)[56], and an ascertainment of whether the lead SNPs and their LD partners (SNPs with an $r^2 > 0.6$ with a lead SNP and no more than 250 kb from it) are protein-altering variants or are associated with cis gene expression in distinct human tissues, among other analyses. The Supplementary Note also highlights the most important results of the bioinformatics analyses and summarizes the conclusions we derive from them.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

GWAS summary statistics can be downloaded from http://www.thessgac.org/data. SNP-level summary statistics from analyses based entirely or in part on 23andMe data can only be reported for up to 10,000 SNPs. For general risk tolerance, we provide association results for all SNPs that passed quality-control filters in a GWAS meta-analysis of general risk tolerance that excludes the research participants from 23andMe; we also provide association results from the complete GWAS (which includes data from 23andMe) for all lead SNPs identified in our discovery GWAS and MTAG analysis of general risk tolerance and for the 4,000 most significant SNPs in the meta-analysis of the discovery and replication GWAS of risk tolerance. For adventurousness, we provide association results from the complete GWAS (which includes only data from 23andMe) for all lead SNPs and for the next 4,000 most significant SNPs. For automobile speeding propensity, drinks per week, ever smoker, number of sexual partners, and the first PC of the four risky behaviors, we provide association results from the complete GWAS for all SNPs that passed quality-control filters. Contact information for the cohorts included in this paper can be found in the Supplementary Note.

## References

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

45. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
46. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
47. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).
48. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
49. Finucane, H. K. et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
50. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. Preprint at https://doi.org/10.1101/145581 (2017).
51. Stringer, S. et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Transl. Psychiatry* **6**, e769 (2016).
52. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
53. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
54. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
55. Buchanan, C. C., Torstenson, E. S., Bush, W. S. & Ritchie, M. D. A comparison of cataloged variation between International HapMap Consortium and 1000 Genomes Project data. *J. Am. Med. Informatics Assoc.* **19**, 289–294 (2012).
56. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

Corresponding author(s): Jonathan P Beauchamp
jonathan.pierre.beauchamp@gmail.com

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

## Experimental design

1. Sample size

Describe how sample size was determined.

The appropriate sample size was determined based on power calculations reported in the preregistered analysis plan on Open Science Framework (https://osf.io/cjx9m/). That information is also reported in Supplementary Note section 2:
"The analysis plan included power calculations assuming that 100,000 individuals in the UKB answered “Yes” (“cases”) to the general-risk-tolerance question, and 270,000 individuals answered “No” (“controls”). Under this assumption, our study would have 73% power to detect single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) of 0.3 and an odds ratio of 1.05 with a genome-wide significance threshold of P = 5×10–8."
Ultimately, for the general risk tolerance phenotype, we combined data from the UKB, from a sample of research participants from 23andMe, and from 10 smaller replication cohorts.

2. Data exclusions

Describe any data exclusions.

Exclusion criteria were specified in the preregistered analysis plan on Open Science Framework. That information is also reported in Supplementary Note section 2:
"The analysis plan instructed all cohorts to limit the analysis to individuals of European ancestry, to exclude individuals with missing covariates, to remove samples that displayed a SNP call rate of less than 95%, and to apply cohort-specific standard quality control filters before imputation."

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Supplementary Information section 5 reports holistic, out-of-sample replication in ~35,000 individuals. Also, we test whether polygenic scores have predictive power in multiple independent prediction cohorts (Supplementary Information section 10). In all cases, the replication record is strong.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not relevant because the study is not experimental.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not relevant because the study is not experimental.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ The statistical test(s) used and whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ Test values indicating whether an effect is present<br>*Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.* |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation) |

*See the web collection on statistics for biologists for further resources and guidance.*

## Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

The URLs section in the main text contains links to software used in this study, and the Supplementary Note describes the software that were used for the various analyses. In summary:
Cohort-level imputation was performed with MaCH/Minimac, BEAGLE, IMPUTE2/4, PBWT, and ShapeIT.
Genetic data was managed with QCtool and BCFtools.
GWAS association was performed with PLINK, SNPtest, Mach2QTL, GCTA, BOLT-LMM, regscan, and R.
Meta-analyses were performed with METAL.
QC of GWAS summary statistics was performed with EasyQC and R.
LD score regressions were done using ldsc v1.0.0 and Python.
Clumping was performed with Plink, 1.90b3p.
Conditional association analysis was performed with GCTA.
Polygenic score weights were generated using LDpred v0.9.09.
Polygenic scores were calculated with PLINK.
SNP heritability was estimated with ldsc v1.0.0, HESS, and GCTA.
Biological annotation and gene-based analyses were conducted using SMR, DEPICT (downloaded Feb 2015), MAGMA v1.06b.
MTAG analyses were conducted using the MTAG software v1.0.1.
Auxiliary statistical analyses were performed with Python, STATA, and R.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. — No eukaryotic cell lines were used

b. Describe the method of cell line authentication used. — No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination. — No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. — No cell lines were used.

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study. — No research animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. — Analyses were conducted on GWAS summary statistics. References to the studies that report covariate-relevant population characteristics are in Supplementary Tables 4-5.