# The Common Genetic Architecture of Anxiety Disorders

Purves, K. L.[1], Coleman, J.R.I.[1,2], Rayner, C.[1], Hettema, J. M.[3], Deckert, J.[4], McIntosh, A. M.[5], Nicodemus, K.[6], Breen, G.[1,2*], Eley, T. C.[1,2*]

[1] King's College London; Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry, Psychology & Neuroscience; London SE5 8AF; UK.

[2] NIHR Biomedical Research Centre for Mental Health; South London and Maudsley NHS Trust; London SE5 8AF; UK.

[3] Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, 23298-0126; USA

[4] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University of Wurzburg, Wurzburg, 97078; Germany

[5] Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, EH4 2XU; UK

[6] Centre for Genomic and Experimental Medicine, MRC Institute of Genetics & Molecular Medicine, The University of Edinburgh, Western General Hospital, Edinburgh, EH4 2XU; UK

Address correspondence to:

Kirstin Purves; Social, Genetic and Developmental Psychiatry Centre; Institute of Psychiatry, Psychology & Neuroscience; King's College London; London SE5 8AF; UK.

kirstin.purves@kcl.ac.uk.

[*] These authors contributed equally to this work

## Abstract

Anxiety disorders are one of the most common, debilitating and costly classes of psychiatric disorders worldwide. Twin studies estimate heritability of anxiety disorders to be between 30% - 60%, depending on specific disorder, age, and level of impairment. Although individual anxiety disorders are considered clinically distinct, they share much of their phenotypic and genetic variance, potentially reflecting an underlying liability distribution. The UK Biobank has collected symptom and disorder level anxiety data on 157,366 individuals across the UK who have contributed their genetic data. We used this dataset to investigate genome-wide associations, SNP based heritability, and genetic correlations in four anxiety phenotypes. These reflect population level current anxiety symptoms as a quantitative phenotype, and three case control phenotypes; severe current anxiety symptoms, probable lifetime generalised anxiety disorder and self-reported lifetime diagnosis of any anxiety disorder. Probable lifetime generalised anxiety disorder and self-reported lifetime diagnosis of any anxiety disorder were meta-analysed with a comparable genome-wide association study of anxiety. Genetic analyses included unrelated Caucasian individuals of Western European ancestry.

Estimates of SNP heritability from common variants ranged between 4% (for population level anxiety symptoms) and 32% (for probable generalised anxiety disorder), and all four UK Biobank anxiety phenotypes are highly genetically correlated. Three genome-wide significant loci were found to be associated with anxiety. Both rs3807866 located in the TMEM106B protein coding region on chromosome 7, and rs2996471 located in the NTRK2 protein coding region on chromosome 9, were associated with self-report of any lifetime anxiety diagnosis. An additional non characterised region on chromosome 9 was associated with both self report of any lifetime anxiety diagnosis (rs10809485), and severe anxiety symptoms (rs17189482). Meta-analysis with a comparable genome-wide association study of anxiety did not result in additional findings. This represents the largest genetic study of anxiety to date - however larger sample sizes will be required to further examine the common genetic architecture underlying anxiety.

## Introduction

Anxiety disorders are amongst the most common classes of psychiatric disorders world-wide (Kessler et al., 2009; Wittchen & Jacobi, 2005). They have a global lifetime prevalence of ~16% (Kessler et al., 2009), and were responsible for a cumulative 24,355 years lost due to disability globally in 2013 (GBD 2013 DALYs and HALE Collaborators et al., 2015). The costs associated with this severe impairment are high –anxiety costs the UK economy £10 billion every year (Fineberg et al., 2013), a cost that is projected to rise to £14 billion by 2026 (McCrone, Dhanasiri, Patel, Knapp, & Lawton-Smith, 2008).

Although several candidate gene studies of anxiety disorders have been carried out over the years, implicating variants including the 5HTTLPR polymorphism of SLC6A4 and the val158met polymorphism of COMT, amongst others, these associations have not proved robust (Smoller, 2016). As has been the case with other complex psychiatric disorders, such as schizophrenia (Ripke et al., 2013) and depression (Major Depressive Disorder Working Group of the PGC, Wray, & Sullivan, 2017), it is likely that environmental influences, in addition to a multitude of common genetic variants with modest effects, are related to anxiety disorders (Smoller, 2016).

Progress has been made towards identifying some of these variants. Several small genome-wide association studies of panic disorder seem to implicate a transmembrane protein variant 132D (TMEM132D), however this gene's function is not yet fully understood (Smoller, 2016). The proportion of variation in generalised anxiety disorder symptoms explained by individual genetic variation (SNP heritability) was estimated at 7.2% in a small sample (n=12282) of Hispanic/Latino adults (Dunn et al., 2017), and 14% in a larger sample of individuals with European ancestry (Otowa et al., 2016). Two genome-wide significant loci were recently found to be associated with anxiety case-ness (rs1709393) and with a quantitative factor score of broad anxiety (rs1067327, within CAMKMT) in a meta-analysis of genome-wide association studies of several anxiety disorders (N= 18000) (Otowa et al., 2016). Finally, a SNP within RBFOX1 (rs13334105) was found to be significantly associated with anxiety sensitivity in a small cohort of twins (Davies et al., 2015). None of these findings have been replicated in independent cohorts. Significantly larger samples than have been available to date will be required to further understand the common genetic architecture of anxiety disorders.

Clinically, anxiety is not thought of as a homogenous disorder, but is divided into several sub-classifications (for example panic disorder, social anxiety disorder, agoraphobia, or specific phobias). Anxiety disorders are moderately heritable. Estimates range from 30-50% with differences seen depending on the participant age, and specific trait or disorder being assessed and aggregate in families (Craske et al., 2017; Polderman et al., 2015).

However, there is evidence to suggest that both the phenotypic (Craske et al., 2009) and genetic (Otowa et al., 2016; Waszczuk, Zavos, Gregory, & Eley, 2014), structure of specific anxiety disorders is broadly shared. Several family and twin studies indicate genetic overlap between the anxiety disorders (Hettema et al., 2005; Roberson-Nay, Eaves, Hettema, Kendler, & Silberg, 2012; Waszczuk et al., 2014), with clear evidence that the shared genetic component between anxiety disorders is larger than the unique contributions to any one disorder (Tambs et al., 2009; Waszczuk et al., 2014). Furthermore, covariance between anxiety disorders and depression is best explained by a single genetic factor (Hettema, Prescott, Myers, Neale, & Kendler, 2005; Waszczuk et al., 2014), with some evidence for additional phobia specific genetic-variance (Hettema et al., 2005; Waszczuk et al., 2014). A recent review summarised a range of research that strongly indicates that current diagnostic distinctions between anxiety disorders are unlikely to reflect biologically distinct disorders (Smoller, 2016).

It may be that this reflects an underlying liability distribution, with variation in anxiety related traits occurring to different degrees across the population, driven in part by multiple common DNA variants (McGrath, Weill, Robinson, Macrae, & Smoller, 2012; Plomin, Haworth, & Davis, 2009). One of a range of possible anxiety disorders may be considered present when some threshold for liability is crossed. If this were the case, it is likely that anxiety disorders share a common polygenic influence, that may explain the shared phenotypic and genetic structure identified in the twin literature.

To explore this possibility, it is important to consider what is shared phenotypically and genetically between anxiety disorders, and anxiety symptomatology across the population. Common across the anxiety disorders is the feature of persistent and inappropriate worry. These features are at the heart of a diagnosis of generalised anxiety disorder (World Health Organization, 1992). As such, it is likely that generalised anxiety symptoms represent the core, non-specific components of anxiety disorders that are likely influenced by the common phenotypic and genetic factors found in previous studies. In support of this idea, a measure of generalised anxiety disorder symptoms has been usefully applied as a screening measure for a wide range of other specific anxiety disorders (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007).

The current study sought to further our understanding of the underlying genetic architecture of anxiety disorders, and anxiety symptoms at the population level. To fully characterise the phenotype, we conducted four genome-wide association analyses in a large community cohort drawn from the UK Biobank (n=157,271). The first phenotype measured total recent generalised anxiety symptoms as a quantitative score across the sample. This measure is intended to reflect population level variation in current anxiety symptoms. The second phenotype includes individuals who surpass a minimum threshold for severity of current generalised anxiety symptoms, and was treated as a case control phenotype. This is

intended to capture individuals who exhibit greater liability towards maladaptive anxiety traits. The third phenotype consisted of individuals who report a life-time professional diagnosis of "any anxiety disorder"; to identify more severe, but non-specific pathological anxiety. The final phenotype identifies a subset of individuals who meet criteria for probable lifetime generalised anxiety disorder, and captures those individuals who have exhibited more severe and specific pathological anxiety.

Four sets of secondary analyses aimed to: (1) identify protein coding genes in the region of any significant genetic loci, (2) estimate the genetic correlations between the four UK Biobank anxiety phenotypes and both the two largest genome wide analyses of anxiety undertaken to date in the Anxiety NeuroGenetics Study (ANGST; Otowa et al. 2016), and other traits more broadly through using LDhub (Zheng et al., 2017), (3) estimate the heritability of the four UK biobank phenotypes, and finally (4) meta-analyse the most comparable UK Biobank anxiety phenotype with the ANGST phenotypes.

# Methods

### Sample and Phenotype Definition

Participants were drawn from a subset of 157,366 genotyped individuals from the UK Biobank, who have completed an online mental health follow-up questionnaire . UK Biobank is a large prospective cohort study providing a resource for investigation of genetic, health and lifestyle determinants of health and illness across the lifespan, and participants range in age between 46 and 80 at the time of completing the mental health questionnaire. Four phenotypes were derived using this data.

**(1)** ***GAD-7 total score.*** This is a measure of recent symptoms of worry and anxiety (n=157,271).

**(2)** ***GAD-7 severity.*** Participants were considered cases if they obtained a total GAD-7 score of 5 or more. This is an indicator of recent anxiety symptoms meeting the minimum threshold for severity as indicated by the GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006) (n = 28,586).

**(3)** ***Probable generalised anxiety disorder (Probable GAD).*** Participants met criteria for likely DSM-IV generalised anxiety disorder based on their responses to the online Composite International Diagnostic Interview (CIDI) Short-form questionnaire (n=26,104).

**(4)** ***Any anxiety disorder.*** Participants met criteria for this phenotype if they reported a professional diagnosis of "any anxiety disorder", and reported ***not*** having a diagnosis of any of the following: schizophrenia; bipolar disorder; autistic spectrum disorder; attention deficit hyperactivity disorder; or eating disorders (n=11,111).

In addition, a control group was selected consisting of a set of screened healthy individuals, who did not meet criteria for any mental health disorder (n = 58,113).

See supplementary materials for more detailed description of the sample, and phenotype creation.

**Genotyping and Quality Control**

Genotype data was collected and processed as part of the UK Biobank extraction and quality control (QC) pipeline (Bycroft et al., 2017). In addition to genotyped data, UK Biobank released an imputed dataset (see Bycroft et al. 2017 for details). Only SNPs that were imputed to the Haplotype Reference Consortium (HRC) reference panel were used for these analyses.

Only SNPs with a minor allele frequency greater than 0.01, and INFO score greater than 0.4 (indicating well imputed variants) were included in analyses. For additional sample and genotyping QC undertaken for these analyses see the supplementary materials.

**Statistical Analyses**

***Genome-Wide Association Analyses.***

Analyses were limited to individuals of European ancestry. Covariates (age, gender, genotyping batch, assessment centre, and the first 6 genetic principal components to account for population stratification) were regressed out of each phenotype using logistic regression. Resulting residuals were used as the dependent variable in four linear genome-wide association analyses, run using BGENIE v1.2 software (Bycroft et al., 2017), testing the association between single common variants and each anxiety phenotype under an additive model. Any variants surpassing a genome-wide significance level of $5x10^{-8}$ were annotated using Region Annotator software (https://github.com/ivankosmos/RegionAnnotator) to identify known protein coding genes in the regions of significance.

**Secondary Analyses**

***SNP heritability and genetic correlations***

Linkage disequilibrium (LD) score regression (Bulik-Sullivan et al., 2015) was used to estimate the proportion of variance explained by common genetic variants for each of the four UKBB phenotypes. Estimates were converted to the liability scale for the three case-control phenotypes ("Any anxiety disorder", "Probable GAD" and "GAD-7 Severity"). For each of these three phenotypes, liability scale conversions were estimated under the following three scenarios: (1) assuming accurate sampling in the UK Biobank (sample

prevalence = population prevalence), (2) assuming under-sampling in the UK Biobank (sample prevalence < population prevalence) and (3) assuming over-sampling in the UK Biobank (sample prevalence > population prevalence). See supplementary materials for sample and population prevalence estimates for each scenario.

LD score regression was also used to estimate the genetic correlation between the four anxiety phenotypes in the UK Biobank and two additional anxiety phenotypes from the Anxiety NeuroGenetics Study (ANGST) (Otowa et al., 2016). The first was a case control phenotype, where a logistic genome-wide association analysis was performed on individuals meeting Composite International Diagnostic Interview (CIDI) criteria for generalised anxiety disorder, social phobia, panic disorder, agoraphobia or specific Phobias (n = 7016) and supernormal controls (n = 14,745). The second was a factor score phenotype, where a single latent factor was derived, and each individual score for this latent factor was entered into a linear genome wide association analysis (n = 18,186).

Exploratory genetic correlations between the four UK Biobank phenotypes and external phenotypes were estimated using LD score regression on LD hub (Zheng et al., 2017). Genetic correlations were reported if they were significant at the Bonferroni corrected threshold of $P < 0.0002$.

***Meta-analysis***

Inverse-variance weighted meta-analysis of the UK Biobank "Any anxiety disorder" and "Probable GAD" phenotypes with the ANGST case control phenotype were performed using the METAL package (Willer, Li, & Abecasis, 2010). No individuals were present in both the UK Biobank and ANGST samples. Total sample in the "Any anxiety disorder" meta-analysis was 101,527, and 87,410 in the "Probable GAD" meta-analysis.

### Code availability

Code for all analyses are available from researchers on request.

## Results

### Genome-Wide Association

The results of the genome-wide associations analyses for the four UK Biobank phenotypes are shown in figures 1-4. Manhattan and Q-Q plots are shown for each phenotype. Table 1 shows the results of Region annotation for regions that surpassed genome-wide significance ($P < 5x10^{-8}$).

***GAD-7 Total Score.*** No SNPs passed the genome-wide significance threshold of $5x10^{-8}$.

***GAD-7 Severity.*** One region on chromosome 9 surpassed the genome-wide significance threshold of $5x10^{-8}$. There were no known protein coding genes in this region, however annotation using RegionAnnotator (https://github.com/ivankosmos/RegionAnnotator) indicated that this region has been significantly associated with neuroticism in an independent genome-wide association. See supplementary materials for region plot for this locus.

***Probable GAD.*** No SNPs passed the genome-wide significance threshold of $5x10^{-8}$.

***Any anxiety disorder.*** Three regions on chromosomes 7 and 9 were significant at the genome-wide threshold of $5x10^{-8}$. The index SNP for the region on chromosome 7 was rs3807866 ($p = 1.1x10^{-8}$). Region annotation indicated the presence of two protein coding genes in this region: Transmembrane Protein 106B (TMEM106B) and Von Willebrand Factor D and EGF Domains (VWDE). Region plots of this locus show that the association is limited to the TMEMB106B gene region (see figure 4). The index SNP for the most significant region on chromosome 9 was rs10809485 ($p=3.3x10^{-12}$). There were no known protein coding genes in this region, however region annotation indicated that this SNP has been significantly associated with neuroticism in a separate genome-wide association (Smith et al., 2016). The index SNP for the second region on chromosome 9 is rs2996471 ($p=7.8x10^{-9}$). Region annotation indicates that this region is in the intron for the protein coding gene Neurotrophic Receptor Tyrosine Kinase 2 (NTRK2). See supplementary material for region plot of these loci.

Figure 1. Manhattan Plot and Q-Q plot for GAD-7 Total Symptom Score genome-wide analysis

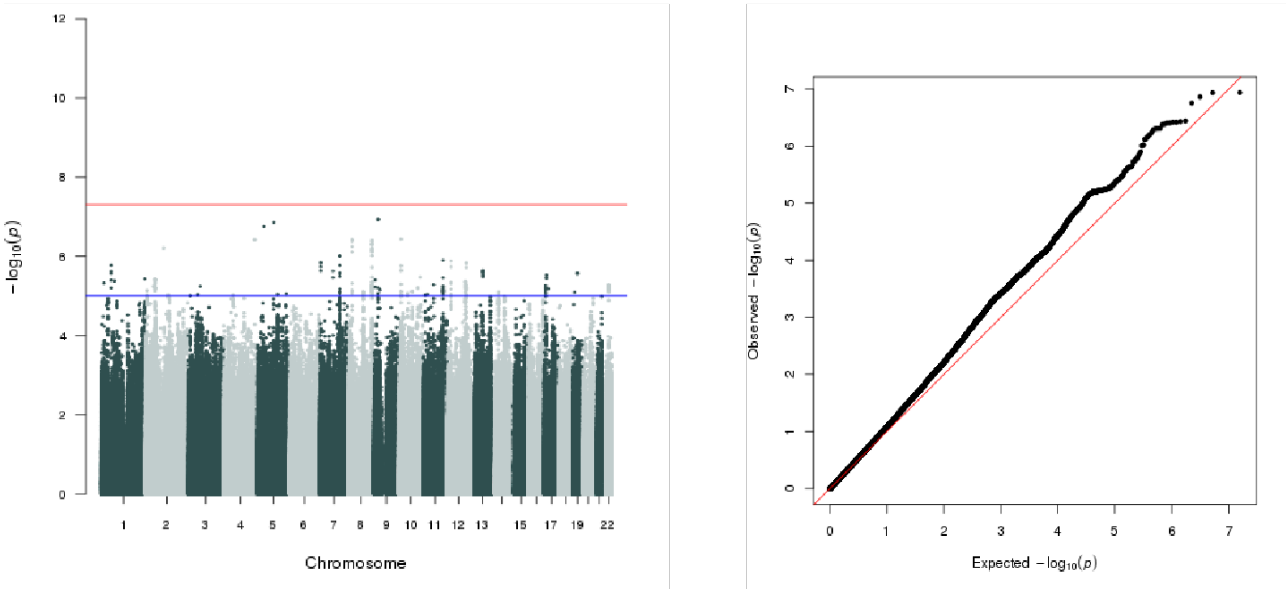

Figure 2. Manhattan Plot and Q-Q plot for GAD-7 Severity Case/Control genome-wide analysis

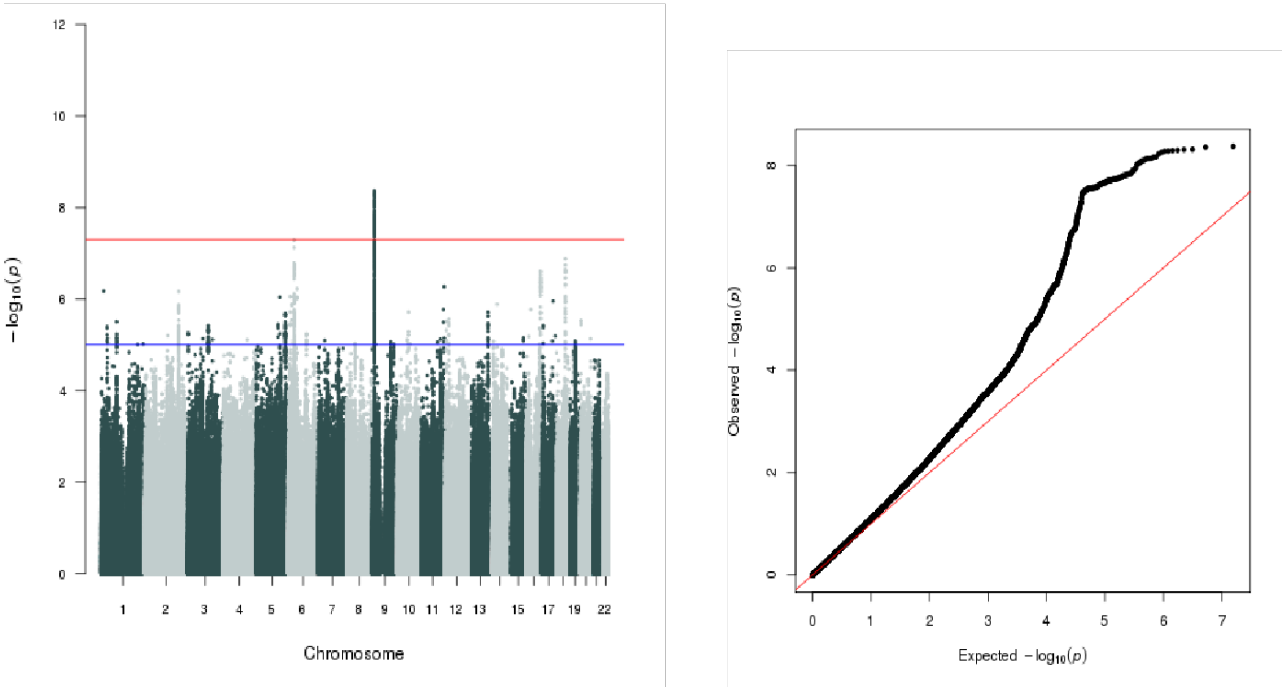


Figure 3. Manhattan Plot and Q-Q plot for Probable GAD Case/Control genome-wide analysis

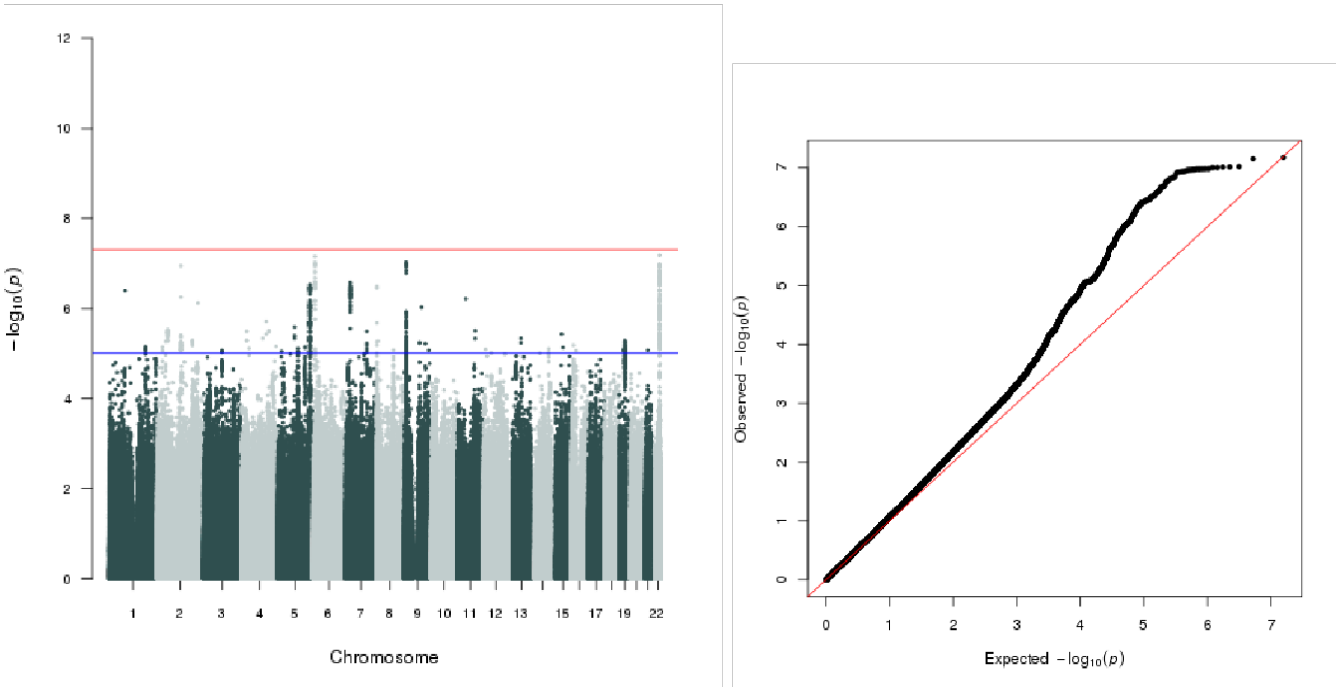

Figure 4. Manhattan Plot and Q-Q plot for Any Anxiety Case/Control genome-wide analysis

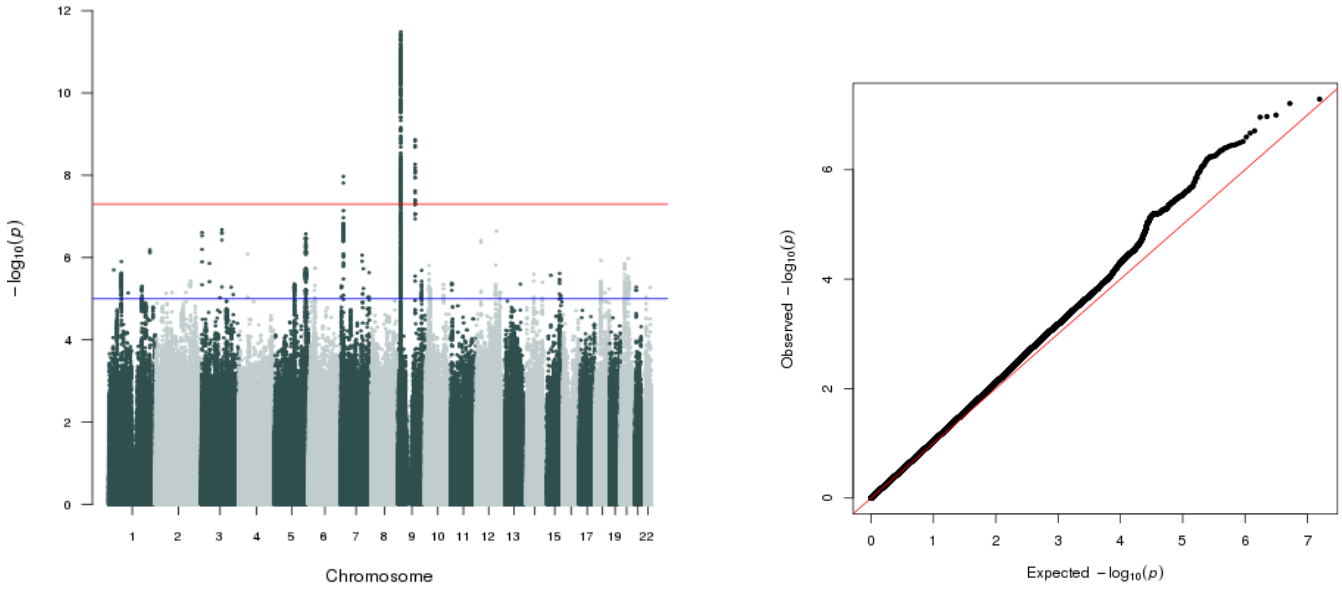


Figure 5. Zoom plot of rs3807866 loci in any anxiety genome-wide analyses

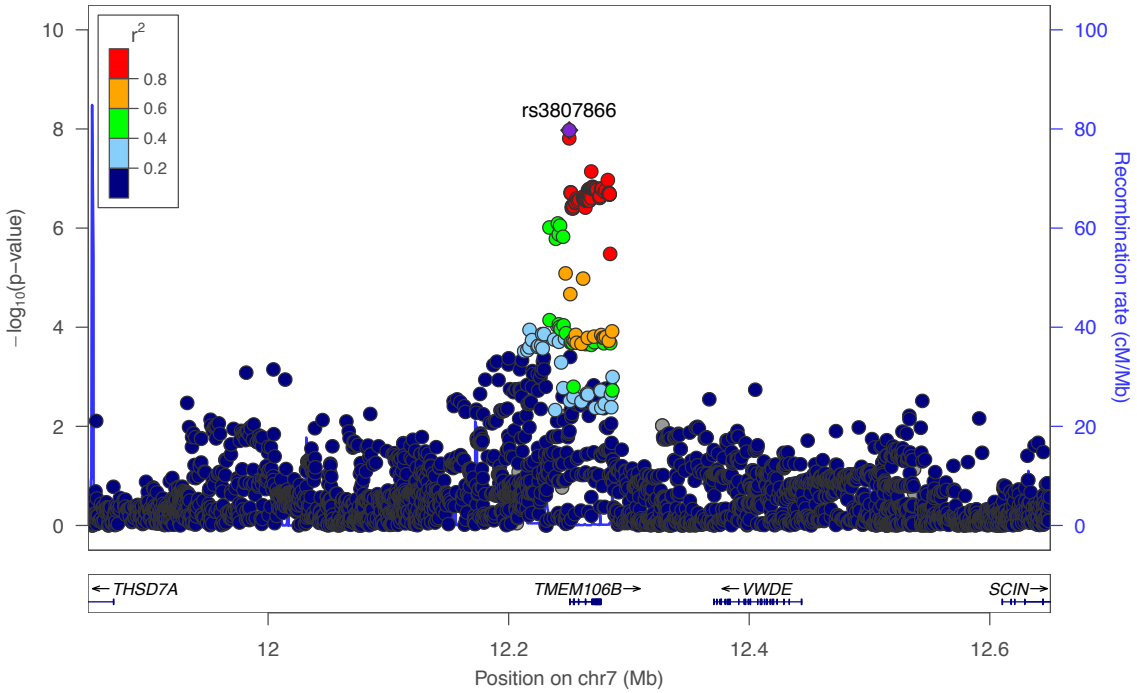

**Table 1. Summary of genome-wide significant associations**

| GWA | Lead SNP ID | Position (hg19) | Protein Coding Genes | Allele | MAF | Beta | P-Value |
|---|---|---|---|---|---|---|---|
| Any Anxiety Disorder Case/Control | rs10809485 | **chr9:** 11,105,574-11,880,898 | -- | G/A | 0.24 | -0.04 | $3.3 \times 10^{-12}$ |
| | rs2996471 | **chr9:** 87,411,744-87,445,656 | NTRK2 | G/T | 0.79 | 0.04 | $7.8 \times 10^{-9}$ |
| | rs3807866 | **chr7:** 12,233,919-12,284,378 | TMEM106B; VWDE | A/G | 0.42 | 0.03 | $1.1 \times 10^{-8}$ |
| GAD-7 Severity Case/Control | rs17189482 | **chr9:** 11,112,537-11,883,299 | -- | G/T | 0.21 | -0.04 | $4.2 \times 10^{-9}$ |

GWA, Genome Wide Analysis; Lead SNP ID, rsid of cardinal SNP in significant region; Position (hg19), Human Genome hg19 position of genome-wide significant region; Allele, Effect Allele/Alternate allele; MAF, Minor allele frequency of effect allele

## SNP Heritability

Table 2 shows the SNP heritability ($h^2$) estimates for the four UK Biobank phenotypes. The observed $h^2$of the "GAD-7 total score" phenotype was 0.044 *(SE = 0.004).* Under the assumption of accurate sampling (i.e. population prevalence rates as seen in this sample), estimates of $h^2$ converted to the liability threshold of "GAD-7 severity" are 0.204 *(SE = 0.015), "*Probable GAD" are 0.324 *(SE = 0.031)*, and "Any anxiety disorder" are 0.216 *(SE = 0.014*

**Table 2. SNP heritability estimates for UKBB anxiety phenotypes**

| Phenotype | Observed $h^2$ | $h^2$ Converted to Liability Scale | | |
|---|---|---|---|---|
| | | *Population Prevalence = Sample Prevalence* | *Population Prevalence = sample prevalence + 10%* | *Population Prevalence = sample prevalence - 10%* |
| GAD-7 Total Score | 0.044 *(0.004)* | --- | --- | --- |
| GAD-7 Severity Case/Control | 0.095 *(0.007)* | 0.204 *(0.015)* | 0.197 *(0.014)* | 0.211 *(0.015)* |
| Probable GAD Case/Control | 0.089 *(0.009)* | 0.324 *(0.031)* | 0.309 *(0.029)* | 0.338 *(0.032)* |
| Any Anxiety Disorder Case/Control | 0.097 *(0.006)* | 0.216 *(0.014)* | 0.209 *(0.014)* | 0.223 *(0.015)* |

$h^2$, SNP heritability estimated using LD regression.

## Genetic Correlation between anxiety phenotypes

Table 3 shows the genetic correlation (rG) between the four UK Biobank anxiety phenotypes and the two ANGST phenotypes. The four UK Biobank phenotypes correlate highly with each other. The highest correlation is between the "GAD-7 total score" and the "GAD-7 Severity" phenotype (rG = 0.92, *se = 0.02*), and the "GAD-7 Severity" phenotype and the "Any anxiety disorder" phenotype (rG = 0.92, *se = 0.03*). The lowest rG is between the "GAD-7 total score" and "Any anxiety disorder" phenotypes (rG = 0.75, *se = 0.04*).

Genetic correlations between the UK Biobank anxiety phenotypes and the ANGST factor score range between 0.52 *(se = 0.16; se = 0.14)* with "GAD-7 total score" and "GAD-7 Severity" respectively, and 0.41 *(se=0.14)* with "Probable GAD". The genetic correlations between the ANGST case control and factor score phenotypes is 0.39 *(se = 0.24)*.

The ANGST case control phenotype has a high genetic correlation with all four UK Biobank anxiety phenotypes. The lowest is with the "GAD-7 total score" (rG = 0.73, *se = 0.17)* and the highest is with the "GAD-7 Severity" (rG = 0.81, *se = 0.17)*. The correlation between ANGST and the most phenotypically comparable UK Biobank case control phenotypes; "Any anxiety disorder" and "Probable GAD", are 0.78 *(se = 0.17)* and 0.79 *(se = 0.19)* respectively.

Table 3. Genetic correlation between UKBB and ANGST anxiety Phenotypes

| | GAD-7 Total Score | GAD-7 Severity Case/Control | Probable GAD Case/Control | Self-Report Anxiety Case/Control | ANGST Anxiety Case/Control | ANGST Anxiety Factor Score |
|---|---|---|---|---|---|---|
| GAD-7 Total Score | --- | | | | | |
| GAD-7 Severity Case/Control | 0.92 *(0.02)* | --- | | | | |
| Probable GAD Case/Control | 0.79 *(0.05)* | 0.91 *(0.04)* | --- | | | |
| Any Anxiety Disorder Case/Control | 0.75 *(0.04)* | 0.92 *(0.03)* | 0.89 *(0.03)* | --- | | |
| ANGST Anxiety Case/Control | 0.73 *(0.17)* | 0.81 *(0.17)* | 0.79 *(0.19)* | 0.78 *(0.17)* | --- | |
| ANGST Anxiety Factor Score | 0.52 *(0.16)* | 0.52 *(0.14)* | 0.41 *(0.14)* | 0.48 *(0.13)* | 0.39 *(0.24)* | --- |

Genetic correlation ; rG *(standard error),* between the four UK Biobank (UKBB) anxiety phenotypes *(shaded)* and Anxiety NeuroGenetics Study (ANGST) case control and factor score anxiety phenotypes (Otowa).

## Genetic Correlation with external phenotypes

Genetic correlations between "Any anxiety disorder" and external traits are presented in table 4. For genetic correlations between the remaining three UK Biobank Anxiety phenotypes and external traits see supplementary tables 2-4. The most significant correlations between "Any anxiety disorder" and external traits were with neuroticism (rG=0.76, *se = 0.04,* p = $3.08 \times 10^{-77}$*)*, depressive symptoms (rG = 0.74, *se = 0.05,* p = $1.14 \times 10^{-49}$), subjective well-being (rG = -0.59, *se = 0.05,* p=$5.9 \times 10^{-29}$) and the Psychiatric Genetics Consortium cross-disorder analysis (rG = 0.47, *se = 0.05,* p = $6.38 \times 10^{-19}$).

## Meta-analysis with ANGST

Due to low genetic correlation, the ANGST factor score was not meta-analysed with any of the UK Biobank anxiety phenotypes.

The ANGST case control phenotype was meta-analysed separately with both of the phenotypically comparable UK Biobank phenotypes ("Any anxiety disorder" and "Probable GAD").

For the "ANGST-Probable GAD"" meta-analysis, No SNPs reached a genome-wide significance threshold of $5x10^{-8}$.

For the "ANGST-Any anxiety disorder" meta-analysis, one region on chromosome 9 was genome-wide significant (cardinal SNP rs10809485; $p=2.55x10^{-12}$). Of note, this is the same index SNP, and the same chromosome 9 region associated with "Any anxiety disorder" in the UK Biobank alone. The effect size of this locus did not differ between the UK Biobank analysis, and the meta-analysis with ANGST (b = 0.04; se = 0.006).

See supplementary figures for Manhattan and Q-Q plots for the meta-analyses with ANGST phenotypes.

**Table 4. Genetic correlation between UKBB self-report anxiety and top associated external traits.**

| External Trait | PMID | Category | rG | se | Z | P-value |
|---|---|---|---|---|---|---|
| **Neuroticism** | 27089181 | personality | 0.76 | 0.04 | -18.60 | $3.08x10^{-77}$ |
| **Depressive symptoms** | 27089181 | psychiatric | 0.74 | 0.05 | -14.82 | $1.14x10^{-49}$ |
| **Subjective well being** | 27089181 | psychiatric | -0.59 | 0.05 | 11.17 | $5.91x10^{-29}$ |
| **PGC cross-disorder analysis** | 23453885 | psychiatric | 0.47 | 0.05 | -8.96 | $3.24x10^{-19}$ |
| **Schizophrenia** | 25056061 | psychiatric | 0.33 | 0.04 | -8.89 | $6.38x10^{-19}$ |
| **Major depressive disorder** | 22472876 | psychiatric | 0.68 | 0.09 | -7.39 | $1.40x10^{-13}$ |
| **Neuroticism** | 24828478 | personality | 0.73 | 0.09 | -7.34 | $2.09x10^{-13}$ |
| **Age of first birth** | 27798627 | reproductive | -0.26 | 0.04 | 6.83 | $8.29x10^{-12}$ |
| **Insomnia** | 28604731 | cognitive | 0.33 | 0.06 | -5.50 | $3.73x10^{-08}$ |
| **Ever vs never smoked** | 20418890 | smoking_behaviour | 0.29 | 0.06 | -4.78 | $1.76x10^{-6}$ |
| **Years of schooling 2016** | 27225129 | education | -0.14 | 0.03 | 4.69 | $2.63x10^{-6}$ |
| **Bipolar disorder** | 21926972 | psychiatric | 0.28 | 0.06 | -4.57 | $4.84x10^{-6}$ |
| **Coronary artery disease** | 26343387 | cardiometabolic | 0.17 | 0.04 | -4.33 | 1.48E-05 |
| **College completion** | 23722424 | Education | -0.20 | 0.05 | 3.83 | 0.0001 |

Genetic correlations between UKBB Self-report anxiety phenotype and external phenotypes generated from Ldhub (ref).
Only traits where correlation exceeded Bonferroni corrected threshold of $p < 0.0002$ shown here.

PMID, PubMed ID of paper for GWA for external trait; rG, Genetic correlation; se, Standard error; Z, Z-score

## Discussion

LD score estimates of SNP heritability for anxiety symptoms at the population level as measured by the GAD-7 was 4.4%. SNP heritability estimates for the "GAD-7 Severity" phenotype, suggestive of greater liability to maladaptive anxiety symptoms, was 20.4%, and estimates for the two severe anxiety phenotypes ("Any anxiety disorder" and "Probable GAD") were 21.6% and 32.4% respectively.

The low SNP heritability estimate of the GAD-7 total score measure may be due to the large number of individuals that score very low on the measure. This floor effect reduces phenotypic variance, thus limiting the ability to detect associated genetic variance.

The SNP heritability estimates for the three categorical anxiety measures ("Any anxiety disorder", "Probable GAD" and "GAD-7 Severity") account for half, or greater than half of additive genetic variation for anxiety disorders as estimated in twin studies (Craske et al., 2017; Polderman et al., 2015). Heritability estimates from twin studies take into account genetic influence of both common and rare genetic variants, whereas SNP heritability estimates only take into account common SNPS. This suggests that a large proportion of heritable variance in severe anxiety is attributable to common genetic variants.

These estimates are much higher than those derived from previous studies of anxiety (Dunn et al., 2017; Otowa et al., 2016), which is likely a reflection of the significantly larger sample size, more homogenous phenotype, and resultant power to detect the influence of common variants in the current study.

The heritability estimate for "Probable GAD" is notably high. It is possible that this is due to the impairment associated with this phenotype. To date, the majority of the twin literature has analysed only anxiety disorder *symptoms*, rather than disorder per se. The latter has a requirement for there being significant associated impairment. One study that specifically explored heritability of anxiety both with and without associated impairment, found preliminary evidence for higher heritability than is usually seen for anxiety (60%) in phenotypes including impairment (Bolton et al., 2006). The current results are in keeping with this possibility which deserves further exploration in other samples.

Genetic correlations between the four UK Biobank phenotypes are uniformly high. This is indicative of a large degree of genetic overlap in the different aspects of anxiety examined. This is not surprising given the findings in the twin literature that there is more variance shared between the anxiety disorders than there is unique to any single disorder (Waszczuk et al., 2014). All four UK Biobank anxiety phenotypes also had a high genetic correlation with an independent case-control anxiety phenotype (Otowa et al., 2016), further supporting this interpretation.

The high genetic correlation between "GAD-7 total score" and the remaining anxiety phenotypes (including the independent Otowa case-control phenotype) indicates that genes associated with the underlying liability towards anxiety symptoms across the population are largely shared with those predisposing individuals to severe or pathological anxiety.

UK Biobank anxiety measures also share a high positive genetic correlation with external measure of internalising traits – including depressive symptoms and neuroticism; and a high negative genetic correlation with subjective well-being. This indicates a shared genetic component between anxiety, and genes predisposing individuals to greater internalising symptomatology, and lower subjective well-being.

No regions of genome-wide significance were found for "GAD-7 total score". Although this quantitative measure allows the use of the whole sample, it should be noted that this phenotype is not normally distributed across the sample. Only 28,586 individuals would score above the minimum threshold for any current symptom severity.

This study finds a region of genome-wide significance on chromosome 9 associated with severity of current generalised anxiety disorder symptoms ("GAD-7 Severity"). This same locus is found to be associated with any anxiety diagnosis, and has been found to be significantly associated with Neuroticism in two previous studies (Okbay et al., 2016; Smith et al., 2016). It is important to note that these were not independent samples. One of these studies was carried out entirely in UK Biobank (Smith et al., 2016) and the other included a small subset of the UK Biobank as a part of their final sample (Okbay et al., 2016).

No regions of genome-wide significance were found for "Probable GAD", despite the high estimate of SNP heritability for the phenotype. This may indicate that this trait is highly polygenic, and is influenced by a great number of variants of smaller effect. Greater sample sizes than the current study will be necessary to detect association with individual variants of small effect.

In addition to the locus on chromosome 9 which was also seen to be associated with "GAD-7 Severity", two genome-wide significant regions were associated with reporting "Any anxiety disorder" diagnosis in the UK Biobank. The locus on chromosome 7 spans approximately 50 Mb and is associated with the protein coding genes Transmembrane Protein 106B (TMEM106B), and Von Willebrand Facor D and EGF Domains (VWDE), however region plotting of this locus indicate that the association is specific to the TMEM106B gene region. TMEM106B is widely expressed throughout normal human cell types and tissue, including the fetal brain, and adult frontal cortex. It encodes for a type II transmembrane protein. Although the function of this protein is unknown, TMEM106B has been associated with frontal lobe temporal dementia (Chen-Plotkin et al., 2012), earlier onset of frontotemporal lobar degeneration (Van Deerlin et al., 2010; van der Zee et al., 2011), particularly in patients with granulin mutations (Cruchaga et al., 2011; Finch et al., 2011), and reduced expression of TMEM106B, seen in combination with enhanced expression of the Progranulin gene, has been associated with pathological process of Alzheimer's Disease (Satoh et al., 2014). Furthermore, this gene was recently found to be associated with both a broad

depression phenotype and probable Major Depressive Disorder in the UK Biobank (Howard et al., 2017).

The second locus, on chromosome 9; (hg19 chr9:87,411,744-87,445,656), is associated with Neurotrophic Receptor Tyrosine Kinase 2 (NTRK2). NTRK2 is a well characterised gene, and is one of the receptors for Brain Derived Neurotrophic Factor (BDNF). NTRK2 has also been implicated in several neuropsychiatric traits and psychiatric disorders including emotional arousal (Spalek et al., 2016), autism (Correia et al., 2010), suicide (Kohli et al., 2010; Murphy et al., 2011), (Torres et al., 2017), Alzheimer's disease (Chen et al., 2006), alcohol dependence (Xu et al., 2007) and treatment response to antidepressants (Dong, Wong, & Licinio, 2009) and mood stabilisers (Wang et al., 2013). It is important to note that many of these associations have been found in candidate gene studies and are less likely to represent robust findings.

This study provides the first well-powered characterisation of the shared genetic architecture of severe and pathological anxiety, however it has several limitations. Analyses were limited to individuals of Caucasian European ancestry in order to minimise the impact of population stratification on the findings. For this reason, it is unlikely that the findings will generalise well to other diverse populations (Martin et al., 2017). The extreme positive skew of the "GAD-7 total score" measure strongly indicates a floor effect for this measure, that likely resulted in less power to detect common variants associated with population level variation in symptom level anxiety. Furthermore, carrying out primary analyses on four phenotypes increases the burden of multiple testing; although it does allow for greater characterisation of the shared genetic contributions to anxiety.

In summary, the present study demonstrated that a large proportion of the broad heritability of severe and pathological anxiety is attributable to common genetic variants. Furthermore, the genetic variance between anxiety diagnoses and symptomatology is largely shared.

## Acknowledgements

This research has been conducted using the UK Biobank Resource, under application 16577.

This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. High

performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

Bibliography

Bolton, D., Eley, T. C., O'Connor, T. G., Perrin, S., Rabe-Hesketh, S., Rijsdijk, F., & Smith, P. (2006). Prevalence and genetic and environmental influences on anxiety disorders in 6-year-old twins. *Psychological Medicine*, *36*(3), 335–344. doi:10.1017/S0033291705006537

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. doi:10.1038/ng.3211

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … Marchini, J. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. *BioRxiv*. doi:10.1101/166298

Chen-Plotkin, A. S., Unger, T. L., Gallagher, M. D., Bill, E., Kwong, L. K., Volpicelli-Daley, L., … Lee, V. M.-Y. (2012). TMEM106B, the risk gene for frontotemporal dementia, is regulated by the microRNA-132/212 cluster and affects progranulin pathways. *The Journal of Neuroscience*, *32*(33), 11213–11227. doi:10.1523/JNEUROSCI.0521-12.2012

Chen, Z.-Y., Jing, D., Bath, K. G., Ieraci, A., Khan, T., Siao, C.-J., … Lee, F. S. (2006). Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science*, *314*(5796), 140–143. doi:10.1126/science.1129663

Correia, C. T., Coutinho, A. M., Sequeira, A. F., Sousa, I. G., Lourenço Venda, L., Almeida, J. P., … Vicente, A. M. (2010). Increased BDNF levels and NTRK2 gene association suggest a disruption of BDNF/TrkB signaling in autism. *Genes, Brain, and Behavior*, *9*(7), 841–848. doi:10.1111/j.1601-183X.2010.00627.x

Craske, M. G., Rauch, S. L., Ursano, R., Prenoveau, J., Pine, D. S., & Zinbarg, R. E. (2009). What is an anxiety disorder? *Depression and Anxiety*, *26*(12), 1066–1085. doi:10.1002/da.20633

Craske, M. G., Stein, M. B., Eley, T. C., Milad, M. R., Holmes, A., Rapee, R. M., & Wittchen, H.-U. (2017). Anxiety disorders. *Nature Reviews. Disease Primers*, *3*, 17024. doi:10.1038/nrdp.2017.24

Cruchaga, C., Graff, C., Chiang, H.-H., Wang, J., Hinrichs, A. L., Spiegel, N., … Goate, A. (2011). Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels. *Archives of Neurology*, *68*(5), 581–586. doi:10.1001/archneurol.2010.350

Davies, M. N., Verdi, S., Burri, A., Trzaskowski, M., Lee, M., Hettema, J. M., … Spector, T. D. (2015). Generalised Anxiety Disorder--A Twin Study of Genetic Architecture, Genome-Wide Association and Differential Gene Expression. *Plos One*, *10*(8), e0134865. doi:10.1371/journal.pone.0134865

Dong, C., Wong, M. L., & Licinio, J. (2009). Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans. *Molecular Psychiatry*, *14*(12), 1105–1118. doi:10.1038/mp.2009.92

Dunn, E. C., Sofer, T., Gallo, L. C., Gogarten, S. M., Kerr, K. F., Chen, C.-Y., … Smoller, J. W. (2017). Genome-wide association study of generalized anxiety symptoms in the Hispanic Community Health Study/Study of Latinos. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *174*(2), 132–143.

doi:10.1002/ajmg.b.32448

Finch, N., Carrasquillo, M. M., Baker, M., Rutherford, N. J., Coppola, G., Dejesus-Hernandez, M., … Rademakers, R. (2011). TMEM106B regulates progranulin levels and the penetrance of FTLD in GRN mutation carriers. *Neurology*, *76*(5), 467–474. doi:10.1212/WNL.0b013e31820a0e3b

Fineberg, N. A., Haddad, P. M., Carpenter, L., Gannon, B., Sharpe, R., Young, A. H., … Sahakian, B. J. (2013). The size, burden and cost of disorders of the brain in the UK. *Journal of Psychopharmacology*, *27*(9), 761–770. doi:10.1177/0269881113495118

GBD 2013 DALYs and HALE Collaborators, Murray, C. J. L., Barber, R. M., Foreman, K. J., Abbasoglu Ozgoren, A., Abd-Allah, F., … et al. (2015). Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. *The Lancet*, *386*(10009), 2145–2191. doi:10.1016/S0140-6736(15)61340-X

Hettema, J. M., Prescott, C. A., Myers, J. M., Neale, M. C., & Kendler, K. S. (2005). The structure of genetic and environmental risk factors for anxiety disorders in men and women. *Archives of General Psychiatry*, *62*(2), 182–189. doi:10.1001/archpsyc.62.2.182

Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., Davies, G., … McIntosh, A. M. (2017). Genome-wide association study of depression phenotypes in UK Biobank (n = 322,580) identifies the enrichment of variants in excitatory synaptic pathways. *BioRxiv*. doi:10.1101/168732

Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., … Wang, P. S. (2009). The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiologia E Psichiatria Sociale*, *18*(1), 23–33.

Kohli, M. A., Salyakina, D., Pfennig, A., Lucae, S., Horstmann, S., Menke, A., … Binder, E. B. (2010). Association of genetic variants in the neurotrophic receptor-encoding gene NTRK2 and a lifetime history of suicide attempts in depressed patients. *Archives of General Psychiatry*, *67*(4), 348–359. doi:10.1001/archgenpsychiatry.2009.201

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–325. doi:10.7326/0003-4819-146-5-200703060-00004

- Major Depressive Disorder Working Group of the PGC, Wray, N. R., & Sullivan, P. F. (2017). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *BioRxiv*. doi:10.1101/167577

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., … Kenny, E. E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *American Journal of Human Genetics*, *100*(4), 635–649. doi:10.1016/j.ajhg.2017.03.004

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). *Paying the price: the cost of mental health care in England to 2026*. London: The Kings Fund.

McGrath, L. M., Weill, S., Robinson, E. B., Macrae, R., & Smoller, J. W. (2012). Bringing a developmental perspective to anxiety genetics. *Development and Psychopathology*, *24*(4), 1179–1193. doi:10.1017/S0954579412000636

Murphy, T. M., Ryan, M., Foster, T., Kelly, C., McClelland, R., O'Grady, J., … Malone, K. M. (2011). Risk and protective genetic variants in suicidal behaviour: association with SLC1A2, SLC1A3, 5-HTR1B &NTRK2 polymorphisms. *Behavioral and Brain Functions*,

*7*, 22. doi:10.1186/1744-9081-7-22

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., … Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624–633. doi:10.1038/ng.3552

Otowa, T., Hek, K., Lee, M., Byrne, E. M., Mirza, S. S., Nivard, M. G., … Hettema, J. M. (2016). Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular Psychiatry*, *21*(10), 1391–1399. doi:10.1038/mp.2015.197

Plomin, R., Haworth, C. M. A., & Davis, O. S. P. (2009). Common disorders are quantitative traits. *Nature Reviews. Genetics*, *10*(12), 872–878. doi:10.1038/nrg2670

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. doi:10.1038/ng.3285

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., … Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, *45*(10), 1150–1159. doi:10.1038/ng.2742

Roberson-Nay, R., Eaves, L. J., Hettema, J. M., Kendler, K. S., & Silberg, J. L. (2012). Childhood separation anxiety disorder and adult onset panic attacks share a common genetic diathesis. *Depression and Anxiety*, *29*(4), 320–327. doi:10.1002/da.21931

Satoh, J.-I., Kino, Y., Kawana, N., Yamamoto, Y., Ishida, T., Saito, Y., & Arima, K. (2014). TMEM106B expression is reduced in Alzheimer's disease brains. *Alzheimer's Research & Therapy*, *6*(2), 17. doi:10.1186/alzrt247

Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E. S., Colodro-Conde, L., Ward, J., … O'Donovan, M. C. (2016). Genome-wide analysis of over 106 000 individuals identifies 9 neuroticism-associated loci. *Molecular Psychiatry*, *21*(6), 749–757. doi:10.1038/mp.2016.49

Smoller, J. W. (2016). The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. *Neuropsychopharmacology*, *41*(1), 297–319. doi:10.1038/npp.2015.266

Spalek, K., Coynel, D., Freytag, V., Hartmann, F., Heck, A., Milnik, A., … Papassotiropoulos, A. (2016). A common NTRK2 variant is associated with emotional arousal and brain white-matter integrity in healthy young subjects. *Translational Psychiatry*, *6*, e758. doi:10.1038/tp.2016.20

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. doi:10.1001/archinte.166.10.1092

Tambs, K., Czajkowsky, N., Røysamb, E., Neale, M. C., Reichborn-Kjennerud, T., Aggen, S. H., … Kendler, K. S. (2009). Structure of genetic and environmental risk factors for dimensional representations of DSM-IV anxiety disorders. *The British Journal of Psychiatry*, *195*(4), 301–307. doi:10.1192/bjp.bp.108.059485

Torres, C. M., Siebert, M., Bock, H., Mota, S. M., Castan, J. U., Scornavacca, F., … Bianchin, M. M. (2017). Tyrosine receptor kinase B gene variants (NTRK2 variants) are associated with depressive disorders in temporal lobe epilepsy. *Epilepsy & Behavior*, *71*(Pt A), 65–72. doi:10.1016/j.yebeh.2017.03.030

Van Deerlin, V. M., Sleiman, P. M. A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N. R., … Lee, V. M.-Y. (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nature Genetics*, *42*(3),

234–239. doi:10.1038/ng.536

van der Zee, J., Van Langenhove, T., Kleinberger, G., Sleegers, K., Engelborghs, S., Vandenberghe, R., … Van Broeckhoven, C. (2011). TMEM106B is associated with frontotemporal lobar degeneration in a clinically diagnosed patient cohort. *Brain: A Journal of Neurology*, *134*(Pt 3), 808–815. doi:10.1093/brain/awr007

Wang, Z., Fan, J., Gao, K., Li, Z., Yi, Z., Wang, L., … Fang, Y. (2013). Neurotrophic tyrosine kinase receptor type 2 (NTRK2) gene associated with treatment response to mood stabilizers in patients with bipolar I disorder. *Journal of Molecular Neuroscience*, *50*(2), 305–310. doi:10.1007/s12031-013-9956-0

Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M., & Eley, T. C. (2014). The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. *JAMA Psychiatry*, *71*(8), 905–916. doi:10.1001/jamapsychiatry.2014.655

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. doi:10.1093/bioinformatics/btq340

Wittchen, H.-U., & Jacobi, F. (2005). Size and burden of mental disorders in Europe--a critical review and appraisal of 27 studies. *European Neuropsychopharmacology*, *15*(4), 357–376. doi:10.1016/j.euroneuro.2005.04.012

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. *Geneva: World Health Organization.*

Xu, K., Anderson, T. R., Neyer, K. M., Lamparella, N., Jenkins, G., Zhou, Z., … Lipsky, R. H. (2007). Nucleotide sequence variation within the human tyrosine kinase B neurotrophin receptor gene: association with antisocial alcohol dependence. *The Pharmacogenomics Journal*, *7*(6), 368–379. doi:10.1038/sj.tpj.6500430

Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., Kemp, J. P., Howe, L., Haycock, P. C., … Neale, B. M. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, *33*(2), 272–279. doi:10.1093/bioinformatics/btw613