Investigation of common, low-frequency and rare genome-wide variation in anorexia nervosa


Anorexia nervosa (AN) is a complex neuropsychiatric disorder presenting with dangerously low body weight, and a deep and persistent fear of gaining weight. To date, only one genome-wide significant locus associated with AN has been identified. We performed an exome-chip based genome-wide association studies (GWAS) in 2158 cases from nine populations of European origin and 15 485 ancestrally matched controls. Unlike previous studies, this GWAS also probed association in low-frequency and rare variants. Sixteen independent variants were taken forward for in silico and de novo replication (11 common and 5 rare). No findings reached genome-wide significance. Two notable common variants were identified: rs10791286, an intronic variant in OPCLM (P = 9.89 × 10−6), and rs7700147, an intergenic variant (P = 2.93 × 10−7). No low-frequency variant associations were identified at genome-wide significance, although the study was well-powered to detect low-frequency variants with large effect sizes, suggesting that there may be no AN loci in this genomic search space with large effect sizes.

Molecular Psychiatry advance online publication, 25 July 2017; doi:10.1038/mp.2017.88

INTRODUCTION

Family studies of anorexia nervosa (AN) have consistently shown that first-degree relatives of AN sufferers have an increased risk of AN, compared with relatives of unaffected individuals.1–4 Twin studies have estimated the heritability of AN at 56%,5 with the majority of remaining variance in liability attributed to non-shared environmental factors (38%).5

Three genome-wide association studies (GWAS) of AN have been conducted to date. The first comprised 1033 AN cases collected as part of the Price Foundation Genetic Study of Anorexia Nervosa (PFS),6 a study of AN families in Italy; 7 the second comprised 1033 AN cases and 15 485 controls; 8 and the third comprised 1033 AN cases from five European countries with 15 485 matched controls.9

In aggregate, these GWAS have failed to identify robust loci for AN, with one notable exception.6 The GWAS performed by the PFS identified a common variant (rs1044793) at 10p11.21 on chromosome 10 (10p11.21) with an odds ratio of 1.30 (odds ratio [OR] = 1.30; 95% confidence interval [CI] = 1.18–1.44; P = 2.93 × 10−5). This variant is associated with greater risk of AN in individuals of European descent, although the study was well-powered to detect low-frequency variants with large effect sizes, suggesting that there may be no AN loci in this genomic search space with large effect sizes.
Anorexia Nervosa and 3733 pediatric controls from the Children’s Hospital of Philadelphia.\(^6\) This study focused on common variation and identified 11 suggestive variants (\(P < 1 \times 10^{-5}\)). None reached genome-wide significance in the primary analysis, although one variant (rs4479806) approached genome-wide significance in an associated secondary analysis. The second study (comprising 2907 cases and 14,860 controls) was carried out by the Genetic Consortium for AN, as part of the Wellcome Trust Case Control Consortium 3 (WTCCC3) effort.\(^7\) This study identified two suggestively associated variants (\(P < 1 \times 10^{-5}\)). Notably, signals at \(P = 4 \times 10^{-6}\), which implies that true signals exist within this data set, but that the study was underpowered for detection. Recently, a third study-meta-analyzed samples from both of these studies, as well as some novel cases, comprising a total of 3495 cases and 10,982 controls. To our knowledge, this study identified the first genome-wide significant locus for AN (index variant rs4622308, \(P = 4.3 \times 10^{-9}\)).\(^8\)

Both previous studies focused on common variation. Here, we conducted, to our knowledge, the first association study that also considered low frequency (minor allele frequency (MAF) \(< 5\%\)) and rare exonic variants in addition to common variation.

**MATERIALS AND METHODS**

**Sample collections**

We conducted a GWAS across nine discovery data sets (the majority overlapping with Genetic Consortium for AN, as part of the Wellcome Trust Case Control Consortium 3 (WTCCC3/WTCCC3 samples), resulting in a total of 2158 cases and 15,485 ancestrally matched controls (Table 1 and Figure 1). All AN cases were female. AN diagnosis was made via semistructured or structured interview, or population assessment strategy using Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria for AN. The amenorrhea criterion was not applied, as this has been shown not to be diagnostically relevant\(^9\) and has since been dropped from DSM-5.\(^10,11\) All cases met criteria for lifetime AN.

Exclusion criteria included confounding medical diagnoses, for example, psychotic conditions, developmental delay or medical or neurological conditions causing weight loss.

Ancestry-matched controls were selected for each AN case set. Both male and female controls were used (Table 1). These were obtained either from existing collaborations, or through genotyping repository (dbGaP) access. Each site obtained ethical approval from the local ethics committee, and all participants provided written informed consent in accordance with the Declaration of Helsinki.

Anorexia Nervosa and 3733 pediatric controls from the Children’s Hospital of Philadelphia.\(^6\) This study focused on common variation and identified 11 suggestive variants (\(P < 1 \times 10^{-5}\)). None reached genome-wide significance in the primary analysis, although one variant (rs4479806) approached genome-wide significance in an associated secondary analysis. The second study (comprising 2907 cases and 14,860 controls) was carried out by the Genetic Consortium for AN, as part of the Wellcome Trust Case Control Consortium 3 (WTCCC3) effort.\(^7\) This study identified two suggestively associated variants (\(P < 1 \times 10^{-5}\)). Notably, signals at \(P = 4 \times 10^{-6}\), which implies that true signals exist within this data set, but that the study was underpowered for detection. Recently, a third study-meta-analyzed samples from both of these studies, as well as some novel cases, comprising a total of 3495 cases and 10,982 controls. To our knowledge, this study identified the first genome-wide significant locus for AN (index variant rs4622308, \(P = 4.3 \times 10^{-9}\)).\(^8\)

**Table 1.** Final numbers of cases and controls after QC

<table>
<thead>
<tr>
<th>Population (abbreviation)</th>
<th>Cases</th>
<th>Controls (% female)</th>
<th>Control source</th>
<th>AN population prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany (DE)</td>
<td>664</td>
<td>2701 (51.8)</td>
<td>Pre-existing study</td>
<td>0.93% (11)</td>
</tr>
<tr>
<td>Finland (FIN)</td>
<td>136</td>
<td>5082 (54.7)</td>
<td>Pre-existing study</td>
<td>2.2% (12)</td>
</tr>
<tr>
<td>France (FR)</td>
<td>215</td>
<td>208 (75.5)</td>
<td>Pre-existing study</td>
<td>0.93% (11)</td>
</tr>
<tr>
<td>Greece (GR)</td>
<td>78</td>
<td>408 (58.1)</td>
<td>In-house</td>
<td>1.3% (13)</td>
</tr>
<tr>
<td>Italy (ITA)</td>
<td>103</td>
<td>48 (100)</td>
<td>In-house</td>
<td>3.0% (15)</td>
</tr>
<tr>
<td>Netherlands (NL)</td>
<td>290</td>
<td>3071 (49.7)</td>
<td>In-house; pre-existing study</td>
<td>1.3% (14)</td>
</tr>
<tr>
<td>Norway (NO)</td>
<td>80</td>
<td>94 (100)</td>
<td>In-house</td>
<td>3.0% (15)</td>
</tr>
<tr>
<td>UK</td>
<td>217</td>
<td>3000 (56.7)</td>
<td>Pre-existing study(^a)</td>
<td>0.64% (16)</td>
</tr>
<tr>
<td>USA</td>
<td>375</td>
<td>873 (50.6)</td>
<td>Obtained from dbGaP</td>
<td>0.9% (17)</td>
</tr>
<tr>
<td>Total</td>
<td>2158</td>
<td>15 485</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AN, anorexia nervosa; QC, quality control.\(^a\)In all, 3000 UK controls were randomly selected from the 9828 samples, genotyped as part of the UK household longitudinal survey. The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analyzed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/.

![Figure 1](https://example.com/f1.png)

**Figure 1.** Geographical distribution of samples across Europe. (a) Distribution of cases across Europe; 375 USA cases are not shown in this diagram. (b) Distribution of controls across Europe; 873 USA controls are not shown in this diagram.
Population prevalence of AN in these populations ranged from 0.4 to 3% (refs 12–18; Table 1).

Genotyping
Cases were genotyped on either the 'Infinium HumanCoreExome-12 BeadChip Kit (Illumina, San Diego, CA, USA),'19 or the 'Infinium HumanCoreExome-24 BeadChip Kit (Illumina),'20 at the Wellcome Trust Sanger Institute. Where possible, controls were selected from existing studies with matching genotyping platforms to cases. Three control cohorts had been genotyped on the ‘Infinium HumanExome-12 BeadChip Kit’ (Table 1). To ameliorate potential confounding due to chip effects, case-control summary statistics across cohort were meta-analyzed using an inverse variance-based test in METAL.28 In order to test the heterogeneity of the results, Cochran’s Q and the I² statistic were computed.

Quality control
Genotypes were called using the GenCall22 and Zcall23 algorithms. At each chip-type quality control (QC) was carried out, and ~14 000 single-nucleotide polymorphisms (SNPs) removed.

Controlling for population stratification
In order to account for population stratification, a principal components analysis was carried out for each cohort separately using the smartpca software.24 Population outliers were identified by merging each population with central European 1000 Genomes data.25 Variance explained by each PC was plotted for each population. In order to be both conservative and consistent across populations, the first 10 principal components were included as covariates in the association testing.

Association testing
Unbalanced case–control ratios can lead to anticonservative P-value estimates.26 This study includes a number of unbalanced strata (Table 1). The likelihood ratio test has been shown to have low type-I error rate allowing a 20 kb window in both directions from the gene. This procedure included 75.44% of variants across 18 118 genes. The European panel of the 1000 Genomes project was used as a reference set to estimate LD between SNPs. The analysis also requires the sample size of the study to be specified; because of the unbalanced nature of the study, the effective sample sizes were given here.

Gene P-values were calculated using MAGMA.41 The top 10% of SNPs per gene were used. Significance was defined using a false discovery rate of 5%.42 There is a risk when assigning SNPs to genes using MAGMA that some highly associated SNP might be assigned to multiple overlapping genes, and this distorts pathway results. SNP–gene assignments were checked for all pathways that reached false discovery rate-corrected significance. No instances of SNPs being assigned to multiple genes were found across these pathways.

Pathway analysis
One of the key motivations of studying complex psychiatric disorders such as AN is the desire to unearth biological pathways underlying disease development. Pathway analysis was performed using summary statistics from the meta-analysis for the full data set.

Four pathway databases were used: the Kyoto Encyclopedia of Genes and Genomes (KEGG),34,35 the Reactome pathway database (REACTOME),36 PANTHER pathway (PANTHER)37,38 and the Gene Ontology database (GO).39,40 These were curated to remove redundancy, resulting in a total of 9083 genes.

The genome-wide significance threshold for burden testing is computed in a similar manner to that for single-point analysis, using Bonferroni correction for the number of genes tested. This results in a genome-wide significance threshold of 5.5 × 10⁻⁵.

Assigning variants to genes
Variants identified associated at P < 1 × 10⁻⁴ were assigned to genes using Ensembl (release 83; Ensembl Genome Browser),29,30 For each variant, all predicted consequences (for example, missense, non-synonymous, and so on) and associated gene transcripts were downloaded and compared. Each variant was associated with only one predicted consequence and one Ensembl gene ID (Ensembl Genome Browser).29

Cluster plot checking
Cluster plots were created for all SNPs reaching P < 1 × 10⁻⁴ in any analysis (cohort-specific or meta-analysis) using ScatterShot.31 SNPs were visually inspected for each cohort, and for cases and controls separately. In instances where multiple cohorts were merged (for example, UK cases), cluster plots were checked separately for each original cohort.

Burden testing
The potential aggregation of rare variants in cases compared with controls was investigated using a gene-based approach. Burden tests were carried out using the Zeggini–Morris burden test32 as implemented in rvtests (Rvtests - Genome Analysis Wiki).

All SNPs with MAF between 0.1 and 5% were included; similar to the single-point analysis, a lower bound of minor allele count ⩽ 5 was used. A list of genes and locations was obtained from the UCSC genome browser (Table Browser: www.genome.ucsc.edu). All genes with at least two qualifying variants in at least two populations were used, resulting in a total of 9083 genes.

Burden tests were carried out for each population individually, and the results meta-analyzed using Stouffer’s method, weighted according to effective sample size.33 The genome-wide significance threshold for burden testing is computed in a similar manner to that for single-point analysis, using Bonferroni correction for the number of genes tested. This results in a genome-wide significance threshold of 5.5 × 10⁻⁵.
Replication

SNPs reaching $P < 1 \times 10^{-4}$ in the discovery stage were prioritized for replication. In total, 16 SNPs were selected.

Replication was carried out using two data sets: one existing in silico data set and one set for de novo genotyping. The in silico data set came from an existing GWAS of AN, genotyped on the Illumina HumanHap610 platform. This data set included 1033 cases and 3733 controls. All cases included in this study were female. Controls were both male and female. The de novo replication cohort consisted of 266 self-volunteered female UK cases, collected through the charity Charlotte’s Helix (www.charlotteshelix.net). All participants were adults and had been diagnosed with AN by their clinician. In addition, all participants completed an online questionnaire based on the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV Module H. The Structured Clinical Interview has been used extensively in epidemiological investigations. The Structured Clinical Interview eating disorder module was modified to capture information on lifetime history of eating disorders including AN, and includes questions on body mass index, age of onset, and experience of eating disorders. DNA from the saliva samples was extracted using standard protocols and was quantified using pico-green. Samples were genotyped on the Infinium HumanExome 12 Beadchip, genotypes were called using GenCall and Zcall algorithms and stringent QC was performed pre- and post-call. In all, 1500 ancestry-matched controls (55% female) were obtained from the UK Household Longitudinal Study.

De novo genotyping was performed using the iPLEX Assay and the MassARRAY System (Agena Bioscience, San Diego, CA, USA) (formerly Sequenom). Sample and SNP QC were carried out within each replication data set, using an 80% sample call rate and a 90% SNP call rate threshold, and a Hardy–Weinberg equilibrium threshold of $10^{-4}$. Five samples and one SNP were removed using these criteria.

Post-QC, 15 SNPs and 261 de novo cases remained. The de novo replication analysis therefore included 15 SNPs, 261 cases and 1500 controls. Genotypes for 12/16 SNPs were available in the in silico replication cohort, across 1033 in silico cases and 3733 controls.

Expression analysis

Gene expression data were obtained from the Genotype-Tissue Expression (GTEx) project web portal, data release version 6 (dbGap Accession phs000424.v6.p1).44–46

Power

The sample sizes used in this study are small in the context of other psychiatric phenotypes. Power to identify genome-wide significant signals was calculated using Quanto.47,48 This study is adequately powered to detect low-frequency alleles with large effect sizes and common alleles with substantial effect sizes (80% power to detect common alleles with odds ratio (OR) > 1.5; low-frequency alleles with OR > 2, Supplementary Figure 1).

Data availability

Genotypes of European cases included in this study are publicly available through the European Genome-Phenome Archive (EGA), under accession number EGAS00001000913, data set EGAD00010001043, with the exception of German and Dutch genotypes. Genotypes for cases from the United States of America may be obtained through dbGaP. Summary statistics are available for download from the PGC website (https://www.med.unc.edu/pgc/results-and-downloads).

RESULTS

GWAS and replication meta-analyses

Association testing was performed separately for each of the nine discovery cohorts within this study (2158 cases, 15 485 controls), and the results were meta-analyzed. No inflation was seen in the QQ plot (Figure 2b). Six variants were identified with $P < 1 \times 10^{-5}$, and nine additional variants with $P < 1 \times 10^{-4}$ (Figure 2a and Supplementary Table 5). Of these, one variant approached genome-wide significance (exm860538/rs199965409, $P = 9.97 \times 10^{-5}$), although this variant is polymorphic only in the Finnish population within these data sets, in the Exome Aggregation Consortium and in the 1000 Genomes project panel data.25 Variants with $P < 1 \times 10^{-4}$ were taken forward for replication.

In total, 16 independent variants were selected for follow-up in one in silico cohort (1033 cases, 3733 controls) and one de novo genotyping cohort (261 cases, 15 000 controls). Of these, five were low frequency (MAF ~ 1%) and 11 were common frequency variants.

Twelve signals passed QC and were polymorphic in the de novo genotyping cohort, of which four were nominally significant (Supplementary Table 6; $P < 0.05$, minimum $P = 0.001$). Eight of twelve SNPs had the same direction of effect as in the discovery GWAS, including three of the four nominally significant variants.

Ten of the sixteen variants were present in the in silico cohort, of which six had the same direction of effect as in the discovery cohort, and one of these six was associated with $P = 0.02$ (Supplementary Table 7).
On the basis of the number of SNPs taken forward for replication, we would not expect to see any variants reaching \( P < 0.05 \) by chance. We also see a higher concordance in direction of effect between discovery and replication cohorts (7/10 in the \textit{in silico} analysis, 8/12 in the \textit{de novo} analysis) than might be expected by chance; however, the number of SNPs tested was too small to achieve statistical significance (\( P = 0.17 \), \( P = 0.19 \), one-sided binomial test).

Five SNPs had the same direction of effect across the meta-analyzed discovery cohort and both replication cohorts. No SNPs reached genome-wide significance in the final global meta-analysis. Two variants were associated with the same direction of effect across discovery and replication cohorts, and reached \( P < 0.05 \) in at least one replication cohort (Table 3).

rs10791286 was associated with risk for AN across all discovery and replication cohorts (Figure 3a, global \( P = 9.89 \times 10^{-6} \), OR 0.84, 95% confidence interval 0.78–0.91). It resides in intron one of the opioid-binding protein/cell adhesion molecule-like (OPCML) gene. Data from the CommonMind Consortium project indicate that this variant is an eQTL for \( \text{OPCML} \) in the dorsolateral prefrontal cortex, and is associated with reduced expression (\( P = 0.014 \) after correction for multiple testing).

\( \text{OPCML} \) has a role in opioid-binding and opioid receptor function and is expressed in a range of neuronal tissues, primarily the cerebellum and cerebellar hemispheres. It has previous associations with body mass index, waist–hip ratio, visceral fat distribution and alcohol dependence, among other phenotypes.

Table 3. Global meta-analysis results

<table>
<thead>
<tr>
<th>Chr</th>
<th>Pos</th>
<th>Id</th>
<th>Associated gene</th>
<th>EA</th>
<th>NEA</th>
<th>EAF</th>
<th>OR</th>
<th>OR_95L</th>
<th>OR_95U</th>
<th>( P )</th>
<th>Het_chisq</th>
<th>N_st (discovery/replication)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>19503281</td>
<td>kgp3754622</td>
<td>—</td>
<td>a</td>
<td>g</td>
<td>0.052</td>
<td>0.81</td>
<td>0.69</td>
<td>0.96</td>
<td>0.016</td>
<td>12.739</td>
<td>8 (6/2)</td>
</tr>
<tr>
<td>10</td>
<td>53754325</td>
<td>rs1904050</td>
<td>PRKG1</td>
<td>a</td>
<td>g</td>
<td>0.20</td>
<td>0.87</td>
<td>0.79</td>
<td>0.95</td>
<td>0.0018</td>
<td>15.033</td>
<td>7 (5/2)</td>
</tr>
<tr>
<td>11</td>
<td>125655014</td>
<td>rs536968</td>
<td>PATE3</td>
<td>a</td>
<td>g</td>
<td>0.12</td>
<td>0.88</td>
<td>0.79</td>
<td>0.98</td>
<td>0.023</td>
<td>12.082</td>
<td>8 (6/2)</td>
</tr>
<tr>
<td>10</td>
<td>122659625</td>
<td>exm860538</td>
<td>WDR11</td>
<td>a</td>
<td>g</td>
<td>&lt;0.01</td>
<td>10.42</td>
<td>4.40</td>
<td>24.69</td>
<td>9.97 \times 10^{-8}</td>
<td>0</td>
<td>1 (1/0)</td>
</tr>
<tr>
<td>4</td>
<td>157167891</td>
<td>rs7700147</td>
<td>ANKRD50</td>
<td>t</td>
<td>c</td>
<td>0.21</td>
<td>1.20</td>
<td>1.10</td>
<td>1.30</td>
<td>2.79 \times 10^{-5}</td>
<td>9.093</td>
<td>8 (6/2)</td>
</tr>
<tr>
<td>3</td>
<td>3482040</td>
<td>exm540361</td>
<td>UHRF1BP1</td>
<td>a</td>
<td>g</td>
<td>&lt;0.01</td>
<td>0.18</td>
<td>0.08</td>
<td>0.37</td>
<td>6.47 \times 10^{-6}</td>
<td>0</td>
<td>1 (1/0)</td>
</tr>
<tr>
<td>6</td>
<td>147840595</td>
<td>rs669830</td>
<td>SAMDS</td>
<td>t</td>
<td>g</td>
<td>0.26</td>
<td>1.11</td>
<td>1.01</td>
<td>1.21</td>
<td>0.029</td>
<td>13.174</td>
<td>5 (3/2)</td>
</tr>
<tr>
<td>7</td>
<td>150748151</td>
<td>rs1703802</td>
<td>CLRN1–AS1</td>
<td>t</td>
<td>g</td>
<td>0.12</td>
<td>0.84</td>
<td>0.75</td>
<td>0.93</td>
<td>0.00085</td>
<td>7.007</td>
<td>8 (6/2)</td>
</tr>
<tr>
<td>17</td>
<td>13082572</td>
<td>exm131068</td>
<td>MYO1D</td>
<td>t</td>
<td>c</td>
<td>0.0011</td>
<td>0.02</td>
<td>0.00</td>
<td>0.10</td>
<td>1.74 \times 10^{-6}</td>
<td>0.276</td>
<td>2 (1/1)</td>
</tr>
<tr>
<td>4</td>
<td>80949829</td>
<td>exm-rs4333130</td>
<td>ANTRX2</td>
<td>t</td>
<td>c</td>
<td>0.38</td>
<td>0.89</td>
<td>0.83</td>
<td>0.94</td>
<td>7.14 \times 10^{-5}</td>
<td>7.526</td>
<td>10 (8/2)</td>
</tr>
<tr>
<td>4</td>
<td>26482021</td>
<td>rs2854030</td>
<td>CCKAR</td>
<td>t</td>
<td>c</td>
<td>0.31</td>
<td>0.88</td>
<td>0.78</td>
<td>0.98</td>
<td>0.021</td>
<td>23.181</td>
<td>8 (6/2)</td>
</tr>
</tbody>
</table>

Abbreviations: CHR, chromosome; EA, effect allele; EAF, effect allele frequency; \( \text{I}_2 \), measure of heterogeneity; NEA, non-effect allele; N_st, number of contributing studies; OR, odds ratio; OR_95L, lower 95% confidence interval; OR_95U, upper 95% confidence interval; \( P \), \( P \)-value; POS, position in hg18. Gene names given are best-predicted consequence from ensembl, where none is available; the nearest gene is given instead in bold.

Figure 3. Odds ratios for two notable single-nucleotide polymorphisms (SNPs) across discovery and replication cohorts. (a) rs10791286 and (b) rs7700147.
The variant itself has no previously reported associations in any phenotype.

rs7700147 was associated with AN across all discovery and replication cohorts (global \(P = 2.93 \times 10^{-5}\), OR 1.2, 95% confidence interval: 1.1, 1.3; Figure 3b). It is an intergenic variant and has no previous associations.

**Burden testing**

Burden testing allows the contribution of multiple low-frequency variants to be aggregated across discrete units (for example, genes). Three genes were identified with \(P < 1 \times 10^{-4}\), although none reached genome-wide significance (Table 4). A further five genes reached \(P < 1 \times 10^{-5}\), but passed inclusion thresholds in one population only (Table 4), and as such are likely to be false-positives.

**FAM96A** has previously been associated with low-density lipoprotein levels and cholesterol\(^{57}\) and is primarily expressed in the liver, lymphocytes and adrenal gland.\(^{44-46}\) KIF7 has no previous phenotype associations and has generally low expression across a wide range of tissues.\(^{44-46}\) C6orf10 has previous associations with visceral fat\(^{55}\) and childhood obesity,\(^{58}\) as well as a number of autoimmune disorders.\(^{59-64}\) C6orf10 is expressed in testes\(^{44-46}\) (see Discussion).

**Biological pathways associated with AN**

Allowing a 20 kb window for SNP to gene assignment identified two pathways significant at \(q < 0.05\): ‘Phospholipase activator’ and ‘GTP-rho binding’ (Table 5).

Using the strictest assignment method of SNPs to genes for the full data set, no pathways were significant after multiple-testing correction. The highest ranking pathway was ‘Calcium ion import’ \((q\text{-value} = 0.069)\).

**DISCUSSION**

To our knowledge, this work constitutes the first examination of low frequency (\(< 1\%\) MAF) and rare exonic variation in AN in the context of a genome-wide scan. No low frequency or rare variant replicating associations were identified, although this study was well-powered to detect low-frequency variants with large effect sizes (Supplementary Figure 1). Although polymorphic only in the Finnish population, rs199965409 approached genome-wide significance. It is a non-synonymous variant with a MAF of 0.5% in the Finnish population.\(^{65,66}\) The variant is within the \(WDR71\) gene, which is associated with hypogonadotropic hypogonadism 14 with or without anosmia.\(^{67-69}\) The clinical features of the disease, such as delayed sexual maturation,\(^{68,70,71}\) suggest that it may be misdiagnosed or comorbid with AN, which may explain its association in the analysis.

Two notable, but common-frequency, signals were identified with consistent direction of effect across discovery and replication cohorts (rs10791286 and rs7700147). These variants had been removed from the first Genetic Consortium for AN, as part of the Wellcome Trust Case Control Consortium 3 (WTCCC3) AN GWAS because of poor cluster plots; therefore, we were not able to compare effect sizes between studies. Burden tests to investigate an aggregation of rare variants within genes rendered three potentially interesting genes, which require further replication.

Studying rare variation presents a range of challenges. The sample sizes required to identify rare variants with modest effect sizes are substantially larger than for common variants. Further, the MAF spectra seen across trans-European populations differ more for rare variants than for common variants, especially when considering genetically distant populations such as Finland and Italy.\(^{25}\) This can reduce the power to detect a signal and achieve replication. There are also many technical challenges to consider when conducting a rare variant study; for example, the inflation seen in association tests at low minor allele count\(^{26}\) and the increased error rate of calling algorithms when applied to rare variants\(^{22,23,72}\). We mitigated against the latter challenge by comprehensively examining cluster plots of >10 000 variants that surpassed a \(P\)-value threshold of \(P < 1 \times 10^{-4}\) in any analysis.

Of the genes potentially implicated through the single-point and burden test analyses, three have associations with metabolic and anthropometric phenotypes (OPCML, C6orf10 and FAM96a). OPCML has previously been associated with waist-to-hip ratio, while C6orf10 has associations with childhood obesity.\(^{58}\) FAM96A has been shown to be associated with metabolic phenotypes such as low-density lipidprotein and cholesterol levels. The associations of these three genes with metabolic and obesity-related phenotypes may indicate some roles for metabolic processes in AN development, although pathway analysis did not corroborate this observation. A growing body of evidence suggests involvement of metabolic processes in AN development, including appetite-satiety pathways, gut motility and gastric-emptying times.\(^{73-79}\) For example, application of the LD Score regression method revealed significant negative genetic correlations between AN and body mass index, insulin, glucose, and lipid phenotypes and significant positive genetic correlations between AN and HDL cholesterol phenotypes.\(^{180}\)

Notably, C6orf10 has been previously associated with childhood obesity.\(^{58}\) This finding is particularly interesting for a number of reasons. First, appetite and satiety dysregulation have been shown to be central to the development of childhood obesity.\(^{81,82}\) In particular, reduced satiety responsiveness (experiencing an urge to eat despite internal ‘full’ signals) and heightened responsiveness to food have a role in increased adiposity. Aberrant responses to satiety signals and reduced responsiveness to food are also operative in AN, suggesting shared biological dysregulation between the two conditions.\(^{83,84}\) Children with increased adiposity are at higher risk of eating disorders\(^{85}\) as they are more likely to engage in high-risk behaviors such as repeated and excessive dieting and erratic or overly rigid eating patterns.\(^{85-89}\) These children are also at higher risk of being bullied about their weight, which may increase weight and shape concerns, body dissatisfaction and a host of related risk factors for AN development.\(^{85-92}\)
The most significant pathway analysis association was with phospholipase activator pathways, which act to catalyze the hydrolysis of glycerophospholipids (GO:0016004 phospholipase activator activity). Phospholipase has a central role in the serotonin-triggered metabolism of arachidonic acid in the brain, which is a common target for antidepressants, such as lithium, carbamazepine (Tegretol), valproate and lamotrigine (Lamictal). These antidepressants have been shown to have varying efficacy in treating AN. Lithium has been used in treatment of AN (with varying success) while carbamazepine and valproate have been successfully used in individuals with complex comorbid eating disorder phenotypes. Finally, lamotrigine has been shown to significantly improve eating disorder and mood symptoms in individuals with binge-eating and purging behaviors.

The second pathway identified as significantly associated with AN was GTP-rho binding. This pathway has a role in brain development, and is regulated by autism-susceptibility candidate gene 2 (AUTS2). This finding is consistent with the comorbidity between AN and autism. Moreover, individuals with AN may be socially withdrawn and exhibit elevated levels of autistic traits associated with lower social functioning. AUTS2 has also been well studied as a candidate gene for alcohol abuse, which is commonly comorbid with eating disorders. There is also a well-established link between GTP-rho activation and cognition. Mice with altered expression of genes regulating Rho-GTPases have been shown to have altered exploratory and anxiety-related behavior, decreased sociability and memory formation, and decreased body weight, among others. These findings are in line with some of the comorbidities and intermediate phenotypes noted in AN, for example, the high comorbidity with anxiety-related disorders.

There is substantial evidence for the involvement of chromatin-modulating genes in the development of autism, schizophrenia and body mass index changes. Given the comorbidity of these disorders with AN, and the potential overlap with autism indicated in the pathway analysis results, we tested for enrichment of chromatin-modulating genes in these results. We obtained a list of 340 genes involved in modifying chromatin accessibility and/or modifying histone marks from existing literature; of these, 30 reached nominal significance in our burden test, substantially more than expected by chance (binomial test, P = 0.0026). Moreover, one of the variants identified in the global meta-analysis (exm540361) lays near a gene included in this list (UHRF1BP1). Together, these results may indicate a role for chromatin-modifying genes in AN, although more work will be needed to investigate this further.

A number of limitations should be borne in mind when evaluating these results. First, the sample size of this study is small. Psychiatric disorders in general require very large sample sizes in order to identify reliable genome-wide significant signals. The current study was powered to detect common variants with substantial OR, and rare variants conferring substantial increases in disease risk (OR > 2). To our knowledge, this was the first time a study has specifically investigated the role of rare variation in AN, and the lack of low-frequency replicating findings may indicate that little advancements may be made in this particular genomic search space.

We did not see any overlap between the pathways identified here and those identified in the recent PGC pathway analysis. However, this may reflect the relatively small sample size of this study, as well as different pathway analysis methodologies used.

In this study we only examined female AN cases of European origin. It has been suggested that the genetics underlying AN development may be easier to assess in an all-male study, as there may be a greater genetic risk required to induce trait expression. The higher relative risk in male subjects may also reflect this. To date, this has not been possible because of the lower prevalence of the disorder in men, resulting in substantially smaller sample sizes. Moreover, if AN is heterogeneous between populations, in order to fully understand the genetic etiology of the disorder, it will be necessary to expand collection to include more diverse samples. Efforts are already underway in a number of Asian populations such as Taiwan, Japan, Korea and China, as well as some South American populations such as Argentina and Brazil.

A caveat to this study is that controls were not screened for AN, and that both male and female controls were used. Given the population prevalence of AN across population of European descent, ~ 80 female and ~ 10 male controls would be expected to have AN diagnoses. Given the low rate of treatment seeking in AN, it would not be possible to confidently screen population-based or previously existing control cohorts for AN.

The underlying biological etiology of AN is complex and has not been elucidated yet. Here we have identified a number of variants that warrant follow-up in larger sample sizes, and which point to a role for metabolic, appetite-related and obesity-related effects, in line with a growing body of evidence for metabolic involvement in AN development. Substantially increased sample sizes and detailed phenotyping to reduce heterogeneity will be necessary to empower the characterization of the genetic architecture of AN.

CONFLICT OF INTEREST
GB has received grant funding and consultancy fees from Eli Lilly, DD is speaker, consultant or on advisory boards of various pharmaceutical companies, including AstraZeneca, Boehringer, Bristol Myers Squibb, Eli Lilly, 28 Genesis Pharma, GlaxoSmithKline, Janssen, Lundbeck, Organon, Sanoﬁ, UniPharma and Wyeth, and he has unrestricted grants from Lilly and AstraZeneca as director of the Sleep Research Unit of Egnition Hospital (National and Kapodistrian University of Athens, Greece). AK is on the Shire Canada BED Advisory Board. JK is a member of SAB of AssurexHealth Inc (unpaid). ML has received lecture honoraria from Lundbeck, AstraZeneca and Biophausia Sweden, and served as scientiﬁc consultant for EPID Research Oy. There exists no other equity ownership, proﬁt-sharing agreements, royalties, or patents. PS is scientiﬁc advisor to Pfizer, Inc. JT received an honorarium for speaking at a diabetic conference for Lilly and royalties from a published book. The remaining authors declare no conﬂicts of interest.

ACKNOWLEDGMENTS
Klarman Family Foundation: This study was supported by a grant from the Klarman Family Foundation.

Wellcome Trust Case Control Consortium: This work was funded by a grant from the WTCCC3 WT088827/Z/09 entitled ‘A genomewide association study of anorexia nervosa’.

Acknowledgments for specific authors: Eleftheria Zeggini is supported by the Wellcome Trust (098051). Lorraine Southam is supported by the Wellcome Trust (098051). Laura Huckins acknowledges Wellcome Trust (098051) and the MRC (MR/J503393/1). Cynthia Bulik acknowledges funding from the Swedish Research Council (VR Dnr: 538–2013–8864). Esther Walton is supported by the German Research Foundation (Wa 3635/1–1). Zeynep Yilmaz is funded by NIH K01MH109782. Allan Kaplan, Robert Levitan and James Kennedy acknowledge support from Ontario Mental Health Foundation for the collection of samples. Jaakko Kaprio is supported by the Academy of Finland (263278).

Charlotte’s Helix: We thank all the probands and parents from the parent-led group, Charlotte’s Helix. This charity was set up by Charlotte Bevan, after her daughter was diagnosed with anorexia nervosa. The charity is deeply committed to supporting biological work, particularly genetics, to help understand anorexia and has set up a database of patients’ names and details. Charlotte’s Helix has been collaborating closely with the King’s College London (KCL) team since 2013 by providing the database of probands, some funding and publicizing the scientific projects through regular blogs on its website and social media.

We gratefully acknowledge the participation of NIHR BRC South London and the Maudsley NHS Foundation Trust (SLaM) BioResource volunteers, and thank the NIHR BRC SLaM BioResource centre and staff for their contribution.

This paper represents independent research partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre and Dementia Unit 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. We gratefully acknowledge capital equipment funding from the Maudsley Charity (Grant Ref. 980) and Guy’s and St Thomas’s Charity (Grant Ref. STR130505).
Low-frequency and rare variation in anorexia nervosa
LM Huikins et al

EATING DISORDERS WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM
RAH Adan 7, 37, L Alfredsson 39, T Ando 40, OA Andreassen 41, 42, T Ando 40, OA Andreassen 41, 42, H Aschauer 43, Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner and Barbara Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn
Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Alzheimer Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania
Roche and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097726, R01-MH-075916, P50MH068981, P50MH084035, R37MH057881 and R37MH057881B1, HHSN27120130031C, AG02219, AG05318 and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Kesuke Hira, Hiyoriyo Toyoyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner and Barbara Bernie Devlin

1University Medical Center Utrecht, Utrecht, The Netherlands; 2Alzheimers Disease Rintvelt, Zeist, The Netherlands; 3Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden; 4Department of Psychosomatic Research, National Institute of Mental Health, National Institute of Neurology and Psychiatry, Tokyo, Japan; 5Division of Mental Health and Addiction, NORMENT, K.G. Jebsen Centre for Psychosis Research, Oslo University Hospital, Oslo, Norway; 6Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 7Biopsychosocial Corporation, Vienna, Austria; 8Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 9Perwelle School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; 10Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; 11Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden; 12INSEEM 1984, Centre of Psychiatry and Neuroscience, Paris, France; 13University of Split School of Medicine, Split, Croatia; 14The Center for Eating Disorders at She appalling, Baltimore, MD, USA; 15Social Genetic and Developmental Psychiatry, King's College London, London, UK; 16NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience, King's College London, SLAM NHS Trust, London, UK; 17Department of Psychiatry and Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 18Karolinska Institutet, Stockholm, Sweden; 19Department of Psychiatry, Medical University of Vienna, Vienna, Austria; 20Clinical Genetics Unit, Department of Women's and Children's Health, University of Padova, Padova, Italy; 21Department of Biomedicine, University Hospital Basel, Basel, Switzerland; 22Department of Genomics, Life and Brain Center, Institute of Human Genetics, University of Bonn, Bonn, Germany; 23Institute of Neurosciences and Medicine, Research Center Jülich, Jülich, Germany; 24School of Psychology, Findor University, Adelaide, SA, Australia; 25Medical Research Council Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK; 26Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, UK; 27Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN, USA; 28CHRU Montpellier, University of Montpellier, Montpellier, France; 29Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA; 30Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 31MIRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 32Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School, Hannover, Germany; 33Department of Dietetics-Nutrition, Harokopio University, Athens, Greece; 34Department of Neurosciences, University of Padova, Padova, Italy; 35Seattle University College of Nursing, Seattle, WA, USA; 36Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; 37First Department of Psychiatry, Athens University Medical School, Athens, Greece; 38University of Nantes, Nantes, France; 39Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA, USA; 40Laboratory of Psychiatric Genetics, Poznan University of Medical Sciences, Poznan, Poland; 41The Barcelona Institute of Science and Technology, Universitat Pompeu Fabra, CIBER Epidemiología y Salud Pública Barcelona, Barcelona, Spain; 42Interdisciplinary Cluster for Applied Genoproteomics-Université de Liège, Liège, Belgium; 43Stanford University, Stanford, CA, USA; 44Broad Institute of MIT and Harvard, Massachusetts General Hospital, MA, USA; 45Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Molecular Psychiatry (2017), 1–12
130Center for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada; 139Department of Public Health, Clinica, University of Helsinki, Helsinki, Finland; 140Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; 141Department of Medicine, Unit of Rheumatology, Karolinska University Hospital, Stockholm, Sweden; 142Department of Psychology, Michigan State University, East Lansing, MI, USA; 143Health Data and Digitalisation, Norwegian Institute of Public Health, Oslo, Norway; 144Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 145Gothenburg University, Gothenburg, Sweden; 146NORMENT—K.G. Jebsen Center for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 147Dr Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 148Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 149Department of Clinical Psychology, American School of Professional Psychology at Argosy University Washington DC, Washington, DC, USA; 150M. Sklodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland; 151Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway; 152K. G. Jebsen Center for Neuropsychiatric Disorders Norway, Oslo, Norway; 153Brain Mind Institute, EPFL, Center for Psychiatric Neuroscience, Department of Psychiatry—CHUV/UNIL, Lausanne, Switzerland; 154Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; 155Department of Psychiatry, University of Naples SUN, Naples, Italy; 156Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 157Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 158QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; 159NORMENT, University of Oslo, Oslo, Norway; 156Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; 159University College Cork, Health Service Executive South, Cork, Ireland; 160MRC Social Genetic and Developmental Psychiatry Centre, Kings College London, London, UK; 161Leiden University Medical Centre, Molecular Epidemiology Section (Department of Medical Statistics), Leiden, The Netherlands; 162Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 163Department of Psychiatry and Behavioral Science, University of North Dakota School of Medicine and Health Sciences, Fargo, ND, USA; 164National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA; 165Boston University School of Medicine, Boston, MA, USA; 166Department of Medicine and Surgery, Section of Neurosciences, University of Salerno, Salerno, Italy; 167University of Naples SUN, Naples, Italy; 168Aarhus University, Denmark; 169Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 170Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; 171Harakopio University, Athens, Greece; 172Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior at the University of California Los Angeles, Los Angeles, CA, USA; 173Kartini Clinic, Portland, OR, USA; 174Institute for Molecular Medicine Finland, Helsinki, Finland; 175Centre de Psychiatrie et Neurosciences—Inserm U894, Paris, France; 176Department of Psychiatry, First Faculty of Medicine, Charles University, Prague, Czech Republic; 177Department of Psychiatry, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 178Institute of Public Health, Ostrava University, Ostrava, Czech Republic; 179Department of Psychiatry, Czech Republic; 180University of Helsinki, Helsinki University Central Hospital, Helsinki, Finland; 181University of Turku, Turku, Finland; 182Department of Child and Adolescent Psychiatry, Poznan University of Medical Sciences, Poznan, Poland; 183Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 184Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 185Department of Genetics, Environment and Mental Health, Norwegian Institute of Public Health, University of Oslo, Oslo, Norway; 186Department of Biometry, University of Helsinki, Helsinki, Finland; 187Section of Eating Disorders, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, UK; 188University of Pisa, Pisa, Italy; 189Ludwig-Maximilians-Universität Munich, Munich, Germany; 190Institute of Psychiatry and Neurology, University of Social Sciences and Humanities, Warsaw, Poland; 191Research Group Clinical Epidemiology, Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany; 192The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 193McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 194Department of Human Biology, J. Craig Venter Institute, La Jolla, CA, USA; 195Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria; 196Department of
REFERENCES


30 Rayner NW, Robertson N, McCarthy MJ. ScatterShot: a Java program for creating cluster plots from Affymetrix and Illumina genotype data. American Society of Human Genetics Meeting (Poster Presentation), American Society of Human Genetics Meeting, Boston Convention & Exhibition Center (BCEC); 22–26 October 2013.


Low-frequency and rare variation in anorexia nervosa
LM Huckins et al.


93 Felder CC, Kanterman RY, Ma AL, Axelrod J. Serotonin stimulates phospholipase A2 and the release of arachidonic acid in hippocampal neurons by a type 2 serotonin receptor that is independent of inositolphospholipid hydrolysis. Proc Natl Acad Sci USA 1990; 87: 2187–2191.


95 Bazinet RP. Is the brain arachidonic acid cascade a common target of drugs used to manage bipolar disorder? Biochem Soc Trans 2009; 37(Pt 5): 1104–1109.


118 Grayson DR, Guidotti A. Merging data from genetic and epigenetic approaches to better understand autistic spectrum disorder. Epigenomics 2016; 8: 85–104.


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2017

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)