Title
Evidence for the role of EPHX2 gene variants in anorexia nervosa.

Permalink
https://escholarship.org/uc/item/1dt527kv

Journal
Molecular psychiatry, 19(6)

ISSN
1476-5578

Authors
Scott-Van Zeeland, AA
Bloss, CS
Tewhey, R
et al.

Publication Date
2014-06-01

DOI
10.1038/mp.2013.91

Peer reviewed
Evidence for the role of EPHX2 gene variants in anorexia nervosa

Anorexia nervosa (AN) and related eating disorders are complex, multifactorial neuropsychiatric conditions with likely rare and common genetic and environmental determinants. To identify genetic variants associated with AN, we pursued a series of sequencing and genotyping studies focusing on the coding regions and upstream sequence of 152 candidate genes in a total of 1205 AN cases and 1948 controls. We identified individual variant associations in the Estrogen Receptor-β (ESR2) gene, as well as a set of rare and common variants in the Epoxide Hydrolase 2 (EPHX2) gene, in an initial sequencing study of 261 early-onset severe AN cases and 73 controls (P = 0.0004). The association of EPHX2 variants was further delineated in: (1) a pooling-based replication study involving an additional 500 AN patients and 500 controls (replication set P = 0.00000016); (2) single-locus studies in a cohort of 386 previously genotyped broadly defined AN cases and 295 female population controls from the Bogalusa Heart Study (BHS) and a cohort of 58 individuals with self-reported eating disturbances and 851 controls (combined smallest single locus P < 0.01). As EPHX2 is known to influence cholesterol metabolism, and AN is often associated with elevated cholesterol levels, we also investigated the association of EPHX2 variants and longitudinal body mass index (BMI) and cholesterol in BHS female and male subjects (N = 229) and found evidence for a modifying effect of a subset of variants on the relationship between cholesterol and BMI (P < 0.01). These findings suggest a novel association of gene variants within EPHX2 to susceptibility to AN and provide a foundation for future study of this important yet poorly understood condition.

Molecular Psychiatry (2014) 19, 724–732; doi:10.1038/mp.2013.91; published online 3 September 2013

Keywords: anorexia nervosa; EPHX2; genomics; hyperlipidemia; pooling; sequencing

INTRODUCTION

A multitude of environmental, behavioral and genetic factors have been shown to be associated with anorexia nervosa (AN) and AN predisposition. However, although AN and its psychological correlates have been shown to be quite heritable (for example, 46–78%;6–9) and have a very high sibling recurrence risk (~10-fold;10), candidate gene and genome-wide searches for common single nucleotide variants (SNVs) and copy number variants (CNVs) that influence AN susceptibility have not yielded statistically compelling and replicable findings to date11–13 except possibly in the case of AN recovery.14

Research focused on the discovery of genes and genetic variants associated with common neuropsychiatric conditions has been greatly aided by the rapid development of genetic technologies, including efficient high-throughput genotyping, CNV detection and next-generation sequencing. For example, recent applications of high-throughput genetic assays have identified unequivocal associations involving rare CNVs with autism and related developmental disorders,15–18 as well as schizophrenia.19–21 However, no genetic studies of other common neuropsychiatric conditions have identified strong associations with either more frequent CNVs or other more abundant forms of genetic variation, such as SNVs and small insertion-deletion variants (indels). This may be attributable to a number of factors including the following: (1) the fact that neuropsychiatric diseases are oligogenic or polygenic in nature, making it difficult to identify each and every variant contributing to the diseases through single-locus analyses; (2) complex diseases such as AN may be influenced by collections of rare SNVs and indels—in addition to more common variant effects—which are hard to detect without large sample sizes, next-generation
sequencing technologies and sophisticated data analytic strategies; and (3) in the absence of deep phenotyping, the broad AN diagnosis may include a very heterogeneous set of patients with unique genetic profiles, confounding the detection of any one gene or set of genes. In light of these issues, the identification of groups of rare variants that collectively contribute to AN that may affect different physiologic pathways or mechanisms will thus require well-characterized and -phenotyped patients, large sample sizes, sophisticated DNA-sequencing strategies or all of these.

We exploited a multistage, large-scale sequencing strategy interrogating the coding regions and neighboring sequence of 152 candidate genes hypothesized to be involved in feeding behaviors, dopamine function, GABA and serotonin signaling, as well as previously identified candidate genes and regions from genome-wide association studies, including OPRD1, CHD9 and EPHX2 (Epoxide Hydrolase 2)\(^9,12,19,20\) to identify genetic variants that contribute to AN. We carried out the sequencing on an initial Discovery cohort (261 AN cases; 73 controls) and a DNA-pooling-based Replication cohort (500 AN cases and 500 controls) to identify associations involving both individual SNVs and indels, as well as collections of rare SNVs and indels, that are associated with AN. In this way, we sought to identify variants that are not likely to be detectable with the current genome-wide association study (GWAS) strategies that focus on a select number of common SNVs that explain some of the heritable components of AN and neuropsychiatric disorders in general.\(^21–23\) We further replicated a subset of the associated variants arising from the two sequencing studies in an additional set of previously genotyped AN and eating disordered (ED) cases (\(N = 444\)) and controls (\(N = 1146\)) using imputation-based methods. Finally, we assessed the impact of these variants on the longitudinal body mass index (BMI) and cholesterol profiles of participants in the Bogalusa Heart Study (BHS),\(^24\) tested for association between EPHX2 variants and relevant psychometric variables, and considered the expression profiles of the associated genes in the human brain via the Allen Brain Atlas.\(^25,26\)

**MATERIALS AND METHODS**

**Samples**

We used DNA samples from 262 individuals diagnosed with AN and 80 controls with extensive phenotype information from the Price Foundation (PF) sample repository for the initial sequencing study.\(^7\) We initially selected a sample of 300 women, self-reported Caucasian, with a clinically diagnosed history of Restricting-type Anorexia or Restricting-type with Purging Anorexia (average age of symptom onset 14 years; see Supplementary Table 1), a lifetime BMI of 15 or less and an assessment of these variants on the longitudinal body mass index (BMI) and cholesterol profiles of participants in the Bogalusa Heart Study (BHS),\(^24\) tested for association between EPHX2 variants and relevant psychometric variables, and considered the expression profiles of the associated genes in the human brain via the Allen Brain Atlas.\(^25,26\)

Ancestry assessment

In order to determine the ancestry of the individuals from the PF repository to be used in the sequencing studies, we exploited a database of 1115 individuals with known ancestries from 11 global populations for whom genotype data are available from the third phase of the International HapMap Project.\(^25\) We used multidimensional scaling analyses on the genotype data with the PF samples along with the individuals with known ancestry. Only individuals who exhibited clear clustering with individuals of European descent were included in the sequencing analysis. In the Discovery stage, one case was excluded because of evidence of cryptic admixture with a Hispanic population, and in the Imputation-based cohort replication, one BHS female and six control females from the Scripps Genomic Health Initiative (SGHI) were excluded because of cryptic admixture.

Target capture

We used the solution-based hybridization target capture technology developed by Agilent\(^25\) to target \(775\) kilobases (kb) of unique DNA sequences covering the exons plus 1 kb upstream from exon 1 for 152 candidate genes involved in feeding behaviors, dopamine, GABA and serotonin signaling, as well as previously identified candidate genes and regions.\(^9,12,19\) Supplementary Table 5 lists all the genes that we studied.

Initial study sequencing and variant calling procedures

After the target capture assays, we performed sequencing with the Illumina GAIIx using indexing across four runs with 12 barcodes per lane. We sequenced all 262 cases and 80 controls. Average coverage for each individual for the targeted regions was 125x, with at least 10x coverage for 93.4% of the targeted regions. To call variants from the sequence data, reads were first aligned to the Human Genome Reference (HG18) using the BFAST alignment program,\(^31\) and the variants were called using CRISP with the default parameters\(^23\) for all targeted regions plus 100 flanking base pairs.

Quality controls

Quality-control steps included exclusion of individuals with greater than 10% missing genotype calls and Hardy-Weinberg equilibrium test \(P < 10^{-4}\). To assess the quality of the variant calls, we included technical control samples (HuRef) on multiple sequencing runs and comparing genotype assignments at the loci for which genotype information was available from a previous sequencing study.\(^36\) These comparisons indicated a high level of fidelity in our sequencing, and variant calling pipelines. We achieved concordance of 99.85% with Sanger sequencing data for the HuRef technical control. Our average between-run concordance for 15 technical replicate samples (inclusive of HuRef) sequenced on multiple runs was 99%. A significant fraction of discordant variants across samples was indels.

Pooling-based replication study sequencing and variant calling procedures

We used the pooling-based sequencing protocol and analysis strategy outlined by Bansal et al.\(^9\) We created 50 total pools, each containing DNA from 20 individuals (25 pools with DNA from 20 AN subjects and 25 pools with DNA from 20 control subjects). In addition, we constructed one pool made up of individuals who had been sequenced previously in the Discovery phase. This allowed us to compare allele frequencies derived from the sequencing of the pool against allele frequencies based on genotype counts from the individual genotype data (see Supplementary Materials) to test the fidelity of the pooling and sequencing processes. A total of 2087 variants were called in this pool. A total of 4 variants were excluded as tri-allelic, 57 variants were missed when compared with the previous genotype information, 17 were called with no support for alternate allele and 15 were from one individual. Allele frequencies estimated from the pooled sequencing for the remaining genotypes showed a very strong correlation with the previously determined genotype-based allele frequencies on the basis of explicit genotype counting (overall \(R^2 = 0.988\); Supplementary Figure 2). The equimolar pools were sequenced on the Illumina HiSeq. Each pool of AN subjects was sequenced to \(\sim 908x\) coverage, or \(\sim 45x\) per individual in the pool and each pool of control subjects was sequenced to \(\sim 895x\) per pool, or 45x per individual. Reads from the pools were aligned to the human reference.
Role of EPHX2 gene variants in AN
AA Scott-Van Zeeland et al

Figure 1. Diagrammatic representation of the study design and workflow for the initial and replication sequencing studies including the number of variants identified and analyzed at each stage. In the Discovery phase, 262 cases with anorexia nervosa (AN) and 80 matched controls were selected for sequencing. Exons and upstream promoter regions for 152 candidate genes were captured for targeted sequencing. A total of 8121 variants were tested for association with AN using both single-locus tests and set-based methods (2380 sets).

Independent replication was pursued in 500 AN cases and 500 controls in the same set of 152 candidate genes (4798 variants; 2380 sets). Additional study cohorts from previous genome-wide association studies (GWAS) were used in a second independent replication analysis among 444 cases and 1146 controls and subphenotype-association tests.

Variation annotations
All variants passing quality-control filters were annotated using a suite of computational and bioinformatics techniques including Polyphen239 and SIFT.40 In addition, identified variants were compared with variants in single nucleotide polymorphism database41 and the 1000 Genomes Database42 to determine their novelty. The results of the functional annotations were used to inform the association analyses as described below.

Statistical analysis in the initial sequencing study
Variants passing quality-control filters were tested for association with AN using simple single-locus analyses based on the regression model-based tests in the software package PLINK.43 In addition, identified variants were compared with variants in single nucleotide polymorphism database41 and the 1000 Genomes Database42 to determine their novelty. The results of the functional annotations were used to inform the association analyses as described below.

Statistical analysis in the pooled DNA-replication study
Variants passing quality-control measures were tested for frequency differences between the pools composed of DNA from AN patients and pools composed of DNA from controls. Allele frequency estimates were obtained by summing DNA-sequencing reads harboring the variant and dividing by the total number of DNA-sequencing reads across the pools mapping to a genomic location covering the variant position. Single-locus tests for frequency differences were pursued with Fisher’s exact test. Set-based tests using the same set derived in the initial sequencing study analysis were conducted by computing the collective frequency of the variants estimated from the sequencing reads in the AN and control pools and comparing these frequencies with Fisher’s exact test after transforming the read counts for the two alleles at each variant to allele counts.37

Imputation-based replication cohorts
For replication, we exploited two previously genotyped groups of cases obtained from different GWAS studies, one that focused on AN specifically (referred to as the ‘GWAS’ cases9) and one with self-reported ED histories obtained from different GWAS studies, one that focused on AN specifically (referred to as the SGHI subjects46).

We then compared genotypes among these cases with two sets of control disease risk with genetic information (referred to as the SGHI subjects46).

variants in each gene and variants in known complexes of genes and pathways. To obtain an accurate estimate of the P-value for the set-based tests, 10,000 permutations of AN/control status were generated, the test was recomputed and the frequency with which a permutation-based test resulted in a value more outlying than the original test statistic was recorded to establish an empirical P-value.

Statistical analysis in the pooled DNA-replication study
Variants passing quality-control measures were tested for frequency differences between the pools composed of DNA from AN patients and pools composed of DNA from controls. Allele frequency estimates were obtained by summing DNA-sequencing reads harboring the variant and dividing by the total number of DNA-sequencing reads across the pools mapping to a genomic location covering the variant position. Single-locus tests for frequency differences were pursued with Fisher’s exact test. Set-based tests using the same set derived in the initial sequencing study analysis were conducted by computing the collective frequency of the variants estimated from the sequencing reads in the AN and control pools and comparing these frequencies with Fisher’s exact test after transforming the read counts for the two alleles at each variant to allele counts.37
individuals having comparable ancestral backgrounds (that is, European) based on genotype profile. One set was derived from the Bogalusa Heart Study and one derived from the SGHI. This comparison included genotyped as well as imputed markers based on 1000 Genomes Phase-I integrated variant set reference haplotypes. We used IMPUTE2 software v2.2.2 (software provided by the Department of Statistics, Oxford University) to impute a subset of associated variants arising from the sequencing studies in these subjects. We used an information threshold of 0.5. For samples that were sequenced and had prior GWAS data, we performed the same imputation to assess concordance for the imputed SNPs. For 128 samples on which we had sequencing and GWAS data, the concordance between sequence-based genotypes and GWAS or imputation-based genotypes for SNPs within EPHX2 was 97.7%. The AN cases used for this second replication had extensive psychometric profiling, and secondary analyses of these phenotypes were pursued.

Longitudinal analyses
As a further set of replications and to explore the phenotypic impact of variants identified from our sequencing studies as associated with AN, we assessed associations between these variants and longitudinal data on BMI and other metabolic phenotypes (for example, cholesterol levels) in the BHS data. The analyses were pursued to investigate the impact of associated variants on the relationship between weight gain (that is, change in BMI over time) and changes in total cholesterol level. A linear mixed model that considered total cholesterol level as the dependent variable and age, degree of European ancestry, BMI, genotype and a BMI × genotype interaction term as independent variables was used for these analyses. Separate analyses for male and female subjects were pursued.

Brain gene-expression analysis
We leveraged the Allen Brain Atlas to explore the expression patterns of genes found to be associated with AN from our sequencing studies. Data for this analysis included EPHX2 expression data from three different donors (H0351.2001, H0351.1009 and H0351.2002) represented by two probes. Ten regions in which both the EPHX2 probes had an absolute Z-score of ≥2.5 and were consistent across more than one donor were noted as over/under expressing the gene. Five of the 13 brain regions are biologically relevant structures and are highlighted in Supplementary Table 4.

RESULTS
Initial sample characteristics
Sample sizes for each stage of data analysis are shown in Table 1, and Supplementary Table 1 further provides information on the phenotypic and clinical characteristics of these individuals. As expected, the Discovery and Replication AN cases are significantly different from the controls with respect to current and previous BMI, anxiety levels and other measures of psychological health (Supplementary Table 1).

Ancestry determination
Multidimensional Scaling analysis with individuals of known ancestries was used to assess homogeneity of our initial sample based on genotype information (see Materials and Methods). Supplementary Figure 1 depicts the multidimensional Scaling results, after excluding one AN individual who demonstrated evidence of admixture, and indicates that the final sample of AN case and control subjects for the Discovery-stage sequencing study cluster with individuals of known European ancestry, suggesting that there is little potential for population stratification and genetic background heterogeneity to influence the association analyses.

Initial sequencing association analysis results
After quality control (see Materials and methods), 7358 SNVs and 763 indels remained for further analysis. We performed single-locus-based association analyses using the logistic regression test implemented in the genetic analysis software package PLINK. No single variant reached genome-wide significance ($P < 10^{-8}$). The two variants with the lowest $P$-values resided in the Estrogen Receptor Beta gene (ESR2; 7.1\times10^{-8}) and had been previously observed and reported in single nucleotide polymorphism database (as rs1256066 and rs944050). We also pursued collapsed set variant analyses using ridge regression methods as described in the Materials and Methods section. A total of 2380 collapsed variant sets were tested for association. Supplementary Table 2 provides a ranked list (by $P$-value from the ridge regression analysis) of the top collapsed sets of variants that fall within the targeted exon regions, the nearest gene and the nature of the collapsing used to define the set (that is, all variants within a targeted exon, predicted functional variants within a region, predicted functional variants within a gene category and so on). The top two sets comprised 35 variants in ITPR3 and 14 variants in EPHX2.

Pooling-based replication analysis results
All variants arising from the pooling-based replication sequencing study were subjected to single-locus and set-based allele frequency difference tests between AN and control pools (see Materials and methods). Of the top ranked sets from the Discovery-stage analysis, only one set of variants in the EPHX2 gene was significantly associated in the pooling-based replication study and thus suggested replication. Eight of the 14 variants initially observed in the EPHX2 gene set (chr8:27456902–27458639) at the Discovery stage were identified in the same set in the Replication stage and are denoted by an asterisk (*) in Table 2. The columns labeled ‘Pools’ in Table 3 provide the frequencies for the EPHX2 gene variants within the replicated set in the pooled replication cohort. We also found evidence for additional replication of the ESR2 gene variants in our other cohorts of previously studied (GWAS) AN cases and self-reported (SGHI) ED women, consistent with the previous evidence for a role for estrogen in AN and prior association with this gene. Note that sets of variants identified in the initial sequencing were tested for association in the pooling-based replication study using combined frequency estimates of the variants defining the sets. In addition, in conducting these tests, only variants that were present in the replication sequencing data were used to form the set (for example, some variants identified in the initial sequencing data were not present in the replication samples).

| Table 1. Summary of study cohorts |
| Sample | N |
|-----------------|---|
| Price foundation phase 1 | |
| Cases | 261 |
| Controls | 73 |
| Price foundation phase 2 | |
| Cases | 500 |
| Controls | 500 |
| Price foundation GWAS | |
| Cases | 386 |
| Scripps Genomic Health Initiative | |
| Cases | 58 |
| Controls | 851 |
| Bogalusa Heart Study | |
| Controls—female | 295 |
| Controls—male | 229 |
| Total cases | 1205 |
| Total controls | 1948 |
study—especially novel variants—were not identified in the replication sequencing study. It can be seen from Supplementary Table 2 that a set of variants in the EPHX2 gene exhibited association in both the initial (14 variant sets; ridge regression \( P = 0.0004 \)) and the replication (8 variant sets; Fisher’s exact test adjusted \( P \)-value after control for multiple comparisons via 10 000 permutations \( = 0.0062 \)) sequencing studies. The EPHX2 gene variants in the set included coding and non-coding variants with the variants that replicated residing primarily in a linkage disequilibrium block that covers the last three exons of the gene, including part of the 3′-untranslated region (UTR) (Table 2).

### Table 2. Significant EPHX2 variants from exon-based set test, discovery and replication

| Replication | Chr | Position | VarType | Ref | Alt | Location | Coding impact | rsID |
|-------------|-----|----------|---------|-----|-----|----------|---------------|-----|
| EPHX2       | 8   | 27457430 | snp      | G   | A   | Intron_15/14 | —             | rs59039594 |
| *           | 8   | 27457059 | snp      | G   | A   | Intron_16/15 | —             | rs2291635  |
| *           | 8   | 27457077 | del      | ACA | —   | Intron_16/15 | —             | —             |
| *           | 8   | 27457179 | snp      | C   | T   | Intron_16/15 | —             | —             |
| *           | 8   | 27457194 | snp      | G   | A   | Exon_17/16 | Synonymous | —             |
| *           | 8   | 27457304 | snp      | G   | A   | Intron_17/16 | —             | —             |
| *           | 8   | 27457623 | snp      | C   | A   | Exon_18/17 | Nonsynonymous | —             |
| *           | 8   | 27457881 | snp      | A   | C   | Exon_19/18 | Synonymous | rs1126452  |
| *           | 8   | 27457991 | snp      | A   | G   | 3UTR      | —             | rs1042032  |
| *           | 8   | 27458049 | snp      | T   | C   | 3UTR      | —             | rs1042064  |
| *           | 8   | 27458411 | snp      | C   | T   | Downstream | —             | rs4149259  |
| *           | 8   | 27458470 | snp      | A   | G   | Downstream | —             | —             |
| *           | 8   | 27458512 | snp      | G   | A   | Downstream | —             | —             |
| *           | 8   | 27458581 | snp      | G   | A   | Downstream | —             | —             |

Abbreviation: EPHX2, Epoxide Hydrolase 2.

### Table 3. Single-locus minor allele frequencies for all cohorts

| Gene    | Chr | Position | rsID          | Discovery Pools | GWAS SGH | Discovery Pools | BHS SGH | Pools P | Combined P |
|---------|-----|----------|---------------|-----------------|----------|----------------|----------|---------|-------------|
| EPHX2   | 8   | 27457059 | rs2291635     | 0.0785          | 0.0859   | 0.0925         | 0.0603   | 0.1319  | 0.1061  | 0.1058 | 0.1122 | 7.45E-02 | 9.18E-03 |
| EPHX2   | 8   | 27457179 | —            | 0.0019          | 0.001    | NA             | 0        | 0       | NA       | NA     | NA     | NA       | NA     |
| EPHX2   | 8   | 27457709 | —            | 0              | 0       | NA             | 0.0068   | 0.002   | NA       | NA     | NA     | NA       | NA     |
| EPHX2   | 8   | 27457881 | rs1126452    | 0.228           | 0.1967   | 0.2487         | 0.2456   | 0.274   | 0.2473  | 0.3034 | 0.2581 | 5.03E-03 | 7.70E-03 |
| EPHX2   | 8   | 27457991 | rs1042032    | 0.228           | 0.1962   | 0.2487         | 0.2456   | 0.2708  | 0.2445  | 0.3034 | 0.2578 | 6.67E-02 | 6.20E-02 |
| EPHX2   | 8   | 27458049 | rs1042064    | 0.228           | 0.2021   | 0.2552         | 0.2411   | 0.2945  | 0.2514  | 0.3102 | 0.2579 | 2.38E-02 | 1.59E-02 |
| EPHX2   | 8   | 27458411 | rs4149259    | 0.1494          | 0.1201   | 0.1554         | 0.1842   | 0.137   | 0.16    | 0.1932 | 0.1410 | 5.93E-03 | 9.11E-02 |
| EPHX2   | 8   | 27458470 | rs189089713  | 0.0019          | 0        | 0              | 0        | 0       | 0       | 0      | 0      | 9.02E-02 | 1.44E-03 |

Abbreviations: BHS, Bogalusa Heart Study; EPHX2, Epoxide Hydrolase 2; ESR, Estrogen Receptor-II; GWAS, genome-wide association study.

Imputation-based single-locus replication

To further test the association between AN and ESR2 and EPHX2 gene variants, we imputed the subset of EPHX2 and ESR2 variants that were implicated from the previous analysis stages in the AN GWAS, BHS and SGH replication cohorts (see Materials and methods). There was some evidence of association with AN status at the single-locus level via a combined analysis, as indicated in Table 3. To further characterize the impact of the associated EPHX2 gene variants, we investigated associations with these variants with measures of depression and anxiety collected on the Discovery AN cases (\( N = 261 \)) (see Supplementary Table 3). Table 4 provides the results of these analyses and suggests that these variants may exhibit association with AN-relevant psychometric traits of Trait Anxiety (TA) and Depression (BDI). Interestingly, we observed an association not only between variants in EPHX2 and BDI and TA but also a significant interaction between EPHX2 gene variants and BMI and BDI scores. As shown in Supplementary Figure 4, women with AN who carry AN-associated EPHX2 variants (rs1126452, rs1042032 and rs1042064) show increasing BDI depression scores with decreasing BMI. No BMI × SNP interaction models for TA were significant, and thus only the main effects are reported.

Longitudinal analyses

Given that the EPHX2 gene is known to influence cholesterol function (as discussed in greater depth in the Discussion), we considered the influence of the subset of associated variants that could be imputed in the BHS data set on longitudinal BMI and cholesterol levels. We found evidence that one EPHX2 gene variant, rs2291635, had an impact on the relationship between weight gain (that is, increase in BMI over time) and cholesterol profile measures (total cholesterol, triglycerides, high-density cholesterol) (Table 4).
lipoprotein cholesterol and low-density lipoprotein cholesterol) based on an interaction model (see Materials and methods and Table 5) among both female (Figure 2) and, interestingly, male subjects from the BHS study. No other EPHX2 gene variants showed significant association with cholesterol profiles.

Gene-expression analysis

Given that EPHX2 has been implicated primarily in cholesterol metabolism with documented expression in tissues such as liver and kidney,51 we performed a post hoc exploration of EPHX2 gene expression using publically available data from a recent meta-analysis by Liang and colleagues (2013)52 and the Allen Brain Atlas (see Materials and methods). Data from a recent exploration of 14 177 expression quantitative trait loci (eQTL) accessed at www.hsph.harvard.edu/liming-liang/software/eqtl/ reveal that one of the replicated SNPs from our study, rs4149259, acts as an eQTL for gene expression of EPHX2 (P = 5.17E-20). Significant EPHX2 gene-expression patterns identified from the Allen Brain Atlas are provided in Supplementary Table 4 and suggest that the EPHX2 gene is expressed in neural tissues of relevance to feeding behaviors, anxiety and other AN-associated phenomena. Namely, elevated expression of EPHX2 observed in the paraventricular nucleus (PVN) of the thalamus is interesting in light of studies linking PVN function to food and water intake,53 weight gain in rats54 and stress response.55 Additionally, abnormal expression in sexually dimorphic regions such as the corpus callosum and the hippocampus may also indicate a sex-specific effect of EPHX2 in the manifestation of AN. Finally, high levels of EPHX2 expression were noted in the subcallosal gyrus, which has been implicated in depression and is functionally connected to the thalamus and limbic system.56

Table 4. EPHX2 psychometric subphenotype analysis in AN

| Variant      | BDI main effecta | BDI BMI × SNPb | Trait anxiety main effectc |
|--------------|------------------|----------------|---------------------------|
|              | β                | P              | β                           | P    | β                  | P                  |
| rs2291635    | 0.266            | 0.458          | −0.162                     | 0.650 | 0.131             | 0.030*             |
| rs1126452    | 1.280            | 0.002*         | −1.180                     | 0.004**| 0.125             | 0.039*             |
| rs1042032    | 1.280            | 0.002**        | −1.180                     | 0.004**| 0.125             | 0.039*             |
| rs1042064    | 1.280            | 0.002**        | −1.180                     | 0.004**| 0.125             | 0.039*             |
| rs4149259    | 1.156            | 0.005          | −1.115                     | 0.007  | 0.046             | 0.449             |

Abbreviations: BDI, Beck Depression Inventory; EPHX2, Epoxide Hydrolase 2.
*P < 0.05; **P < 0.005; 10N = 257; 10N = 254.

Table 5. EPHX2 rs2291635 associations with cholesterol profile

|          | Females | Males |
|----------|---------|-------|
|          | Total Cholesterol | Triglycerides | HDL Cholesterol | LDL Cholesterol |
|          | BMI | SNP | BMI × SNP | BMI | SNP | BMI × SNP |
| β | P | β | P | β | P | β | P |
| 1.03 | 8.5 × 10^{-12} | 12.69 | 0.045* | 0.636 | 0.009** | 1.22 | 1.43 × 10^{-8} | 13.28 | 0.107 | 0.577 | 0.044* |
| 3.36 | 4.15 × 10^{-33} | 20.06 | 0.068 | 1.11 | 0.014* | 5.80 | 5.90 × 10^{-28} | 20.31 | 0.290 | 0.920 | 0.208 |
| −0.054 | 3.08 × 10^{-13} | 7.39 | 0.031* | 0.188 | 0.179 | −0.97 | 7.60 × 10^{-19} | 4.42 | 0.273 | 0.228 | 0.138 |
| 0.946 | 4.18 × 10^{-12} | 0.750 | 0.896 | 0.204 | 0.352 | 1.31 | 2.62 × 10^{-12} | 15.13 | 0.034 | 0.650 | 0.009** |

Abbreviations: BMI, body mass index; EPHX2, Epoxide Hydrolase 2; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SNP, single-nucleotide length polymorphism.
*P < 0.05; **P < 0.01.

DISCUSSION

AN is likely influenced by complex interactions between genetic variants, environmental and social factors. Whereas some genetic variants that contribute to these interactions are likely common, results of the only GWAS analyses to date did not show strong associations between common variants and AN,9 although it was limited by small sample size. We therefore leveraged DNA-sequencing studies in a multistage design to identify and replicate both common and rare variants in biologically relevant AN candidate genes that would not necessarily have been found in common variant-based candidate gene and GWAS analyses. Our initial sequencing study focused on a well-characterized cohort of 261 early-onset AN-affected individuals (mean 14 years) and 73 controls followed by a pooling-based replication study involving a much larger set of (n = 500) AN cases and (n = 500) controls (mean onset 16 years). We searched for both single-locus associations and collapsed rare variant set-based associations in
Role of EPHX2 gene variants in AN
AA Scott-Van Zeeland et al

both initial and replication stages. We found some evidence that ESR2 may harbor variants associated with AN, which is interesting in light of the fact that AN is observed in female subjects to a much higher degree than in male subjects, and given its typical onset in adolescence.57 This finding is also consistent with previous studies suggesting that estrogen and estrogen receptors may have a role in mediating the disease.49,58

Importantly, we found more statistically compelling evidence for set-based associations involving a collection of rare variants in the EPHX2 gene and were able to replicate a subset of these variants in additional case cohorts and controls. EPHX2 was included in the gene list for targeted sequencing because of its presence in a list of top associated genes from a preliminary GWAS analysis.58 Still, the role of EPHX2 in AN was not obvious, so we explored potential links between EPHX2 function and AN. The EPHX2 gene is expressed in vascular and non-vascular neural tissues of relevance to AN and is known to influence cholesterol function,51,59 suggesting multiple potential roles in the etiology or maintenance of AN. Interestingly, it has been well documented that a substantial proportion of malnourished individuals with AN have high serum cholesterol,60 and this may be due to genetic variation.61 In fact, the EPHX2 gene variants were also found to be associated with total cholesterol levels measured on one of our replication cohorts (BHS) as well as comorbid symptoms that are common in AN for a large proportion of our study subjects.

The EPHX2 gene codes for the soluble epoxide hydrolase (sEH) protein, which is widely expressed in a variety of tissues, including the liver, lungs, kidney, heart, brain, adrenals, spleen, intestines, urinary bladder, placenta, skin, mammary gland, testis, leukocytes, vascular endothelium and smooth muscle but is most highly expressed in the liver and kidney.51,62,63 sEH catalyzes the hydrolysis of epoxides to their corresponding diols, easily metabolizes both saturated and unsaturated epoxy fatty acids and is highly conserved with nearly all functioning polymorphisms having been defined.51 The specific activity of sEH varies 500-fold in the human liver, suggesting a wide range of potential regulatory mechanisms,64 and sEH gene transcription may also be induced by sex hormones and regulated by the hypothalamic–pituitary–gonadal axis.51 Through its complex epoxide hydrolisys of epoxy eicosatetraenoic acids (EETs) into diols (DHETs), sEH reduces the number of EETs ready for release by phospholipases and may stop the biological activity of these lipids.62 Together, they influence the regulation of inflammation, blood pressure, lipid and carbohydrate metabolism.51

Relevant to this study are previous animal studies that reveal a dynamic change in sEH activity in response to diet. Mice fed high-carbohydrate high-fat (HCHF) diets show typical sequelae of metabolic syndrome, including increased body weight, abdominal fat and plasma lipid abnormalities, among others. In these obese animals, sEH activity in liver was 18% higher than mice fed a standard diet.65 However, when rats fed a HCHF diet were given an sEH inhibitor, metabolic, liver and cardiovascular abnormalities were attenuated, suggesting a direct role for sEH in the etiology of dyslipidemia in response to diet. Total sEH activity was also found to be higher in adipose tissue of rats fed a HCHF diet.66 These studies, together with the findings of this study, may suggest a role for EPHX2 and the gene product sEH in lipid regulation in response to diet.

The observation that AN patients often display hypercholesterolemia in addition to their condition is counterintuitive, given the under-nutrition and low body weight of affected individuals. Fortunately, recovery from anorexia nearly always leads to recovery from hypercholesterolemia, even in very severe cases.67 It has been hypothesized that low levels of cholesterol may decrease the activity of serotonin receptors and transporters and that significantly lower cholesterol levels are associated with depressive symptoms, impulsive/self-harmful behavior (cutting and/or burning) and suicide thoughts/attempts in anorexia patients.67 Moreover, lower cholesterol levels have been associated with increased suicidality more broadly, including ideation and attempts, in depressed patients.68 Notably, from this research, it has been suggested that among depressed patients, BMI and total cholesterol had a negative correlation, and patients with higher cholesterol levels were observed to be significantly less depressed, impulsive and suicidal. We found evidence that a subset of EPHX2 gene variants associated with AN also influence the relationship between increases in BMI and total cholesterol. These data may suggest a role for EPHX2 gene variants in mediating AN-associated physiological changes consistent with previous studies showing an association between genetic variants within EPHX2 and association with lipid profiles and cardiovascular disease.49,59

The observed association between AN and variants in EPHX2 is particularly interesting in light of the EPHX2 gene-expression pattern observed in the brain. Data from the Allen Brain Atlas indicate EPHX2 enrichment in tissues associated with feeding behaviors, depression and stress response, including the paraventricular nucleus and subcallosal gyrus. Although our investigation of EPHX2 gene expression in the brain in this study was exploratory, future studies leveraging functional or structural brain imaging may clarify neural correlates of EPHX2 genetic variants and contribution to AN-related phenotypes. For example, previous neuroimaging genetic studies have identified associations between genetic variants in the serotonin transporter, anxiety and neural tissues of relevance such as amygdala and cingulate,67 and correlated CNTNAP2 gene-expression patterns in brain to frontostriatal functional connectivity patterns associated with variants in the CNTNAP2 gene.72

It will be important to verify the significance of the variants we have identified in a number of ways.73 First, replication of the variants we have identified in larger data sets is crucial. Second, identifying additional phenotypes that may be influenced by the variants we found to be associated with AN would help put in perspective the physiological and neural pathways these genetic variants influence to impact AN susceptibility. For example, it would be interesting to evaluate associations between cholesterol-related phenotypes and the EPHX2 gene in a sample of individuals with AN. Third, it would be important to assess the functional significance of the variants in order to characterize the molecular pathways that they influence, specifically whether they are associated with lower basal sEH activity. Fourth, imaging-genetics studies that evaluate possible links between EPHX2 and structural and functional measures of the brain regions implicated in AN may provide further insight into the relationship between this gene and AN susceptibility. Taken together, our study represents the largest sequencing study of AN to date, and we hope that it will set the stage for further work into this debilitating and life-threatening disease.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
This study was made possible by a generous gift from the Price Foundation. AVZ, CSB, RT, V8, AT, AC, PP, TP, NV, RT, GZ, SL, SM and NJS are all supported in part by NIH grant 5 U01 DA024417; and Scripps Genomic Medicine. NJS and his laboratory are also supported in part by NIH grants: 5 U01 DA024417, 5 R01 HL089655, 5 R01 DA030976, 5 R01 AG035920, 1 R01 MH93500, 2 U19 A063603, 2 U19 AG023122, 5 P01 AG017734 as well as the Stand Up To Cancer Foundation. M Strober receives support from the Resnick Chair in Eating Disorders. B Shih is in part supported by SK01DK087813-02, and CSB is in part supported by NIH grant 1R21HG005747-01. The Price Foundation Collaborative Group has been responsible for the data collection, curation, management and oversight of the DNA samples used in this report. The members of this group have also provided feedback on this and other reports making use of the samples.

Molecular Psychiatry (2014), 724 – 732 © 2014 Macmillan Publishers Limited
REFERENCES

1 Dalle Grave R. Eating disorders: progress and challenges. Eur J Intern Med 2011; 22: 153–160.

2 Kaye WH, Fudge JL, Paulus M. New insights into symptoms and neurocircuitry function of anorexia nervosa. Nat Rev Neurosci 2009; 10: 573–584.

3 Wade TD, Bulik CM, Neale M, Kendall KS. Anorexia nervosa and major depression: shared genetic and environmental risk factors. Am J Psychiatry 2000; 157: 469–471.

4 Klump KL, Miller KB, Keel PK, McGuire M, Iacono WG. Genetic and environmental influences on anorexia nervosa syndromes in a population-based twin sample. Psychol Med 2001; 31: 737–740.

5 Kortegaard LS, Hoerder K, Joergensen J, Gillberg C, Kyvik KO. A preliminary association study on common SNPs and rare CNVs in anorexia nervosa. Mol Psychiatry 2011; 16: 949–959.

6 Bulik CM, Sullivan PF, Tozzi F, Farge B, Lichtenstein P, Pedersen NL. Prevalence, heritability, and prospective risk factors for anorexia nervosa. Arch Gen Psychiatry 2006; 63: 305–312.

7 Lilenfeld LR, Kaye WH, Greeno CG, Merikangas KR, Plotnicov K, Pollice C et al. A family controlled study of anorexia nervosa and bulimia nervosa: psychiatric disorders in first-degree relatives and effects of proband comorbidity. Am J Psychiatry 1998; 55: 603–610.

8 Strober M, Freeman R, Lampert C, Diamond J, Kaye W. Controlled study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of parallel syndromes. Am J Psychiatry 2000; 157: 393–401.

9 Wang K, Zhang H, Bress SS, Duvuuri V, Kaye W, Schork NJ et al. A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. Mol Psychiatry 2011; 16: 949–959.

10 Grice DE, Halmi KA, Fichter MM, Strober M, Woodside DB, Treasure JT et al. The Temperament and Character Inventory (TCI): A Guide to its use and Development. Washington University: St Louis, MO, USA, 1994.

11 Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways. Mol Psychiatry 2011; 16: 949–959.

12 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM et al. Rare structural variants disrupt multiple genes in neurodevelopmental phenotypes. PLoS Genet 2011; 7: e1002334.

13 Hoerder K, Joergensen J, Gillberg C, Kyvik KO. A preliminary association study on common SNPs and rare CNVs in anorexia nervosa. Mol Psychiatry 2011; 16: 949–959.

14 Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C et al. A copy number variation morbidity map of developmental delay. Nat Genet 2011; 43: 838–846.

15 Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikedu M, Ruderfer D et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012; 17: 142–153.

16 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008; 320: 539–543.

17 McClellan SE, Makarov V, Kirov G, Addington K, McClellan J, Yoon S et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 2009; 41: 1223–1227.

18 Vacc V, McCarthy S, Malhotra D, Murray F, Chou HH, Peoples A et al. Duplication of the neuregulin receptor gene NR3E3 confer significant risk for schizophrenia. Nature 2011; 471: 499–503.

19 Root TL, Szatmavicz JP, Jonassaint CR, Thornton LM, Pinheiro AP, Strober M et al. Association of candidate genes with phenotypic traits relevant to anorexia nervosa. Eur Eat Disord Rev 2011; 19: 487–493.

20 Pinheiro AP, Bulik CM, Thornton LM, Sullivan PF, Root TL, Bress SS et al. Association study of 182 candidate genes in anorexia nervosa. Am J Med Genet B Neuropsychiatr Genet 2010; 153B: 1070–1080.

21 Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 2009; 19: 212–219.

22 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ et al. Finding the missing heritability of complex diseases. Nature 2009; 461: 747–753.

23 Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009; 10: 241–251.

24 Smith EN, Chen W, Kahonen M, Kettunen J, Lehtimaki T, Penelton L et al. Large-itudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. PLoS Genet 2010; 6: e1001094.

25 Jones AR, Overly CC, Sunkim SM. The Allen Brain Atlas: 5 years and beyond. Nat Rev Neurosci 2009; 10: 821–828.

26 Hawrylycz M, Ng L, Page D, Morris J, Lau C, Faber S et al. Multi-scale correlation structure of gene expression in the brain. Neural Netw 2011; 24: 933–942.

27 Cloninger CR. The Temperament and Character Inventory (TCI): A Guide to its Development and Use. Washington University: St Louis, MO, USA, 1994.

28 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961; 4: 561–571.
59 Sura P, Sura R, Enayetallah AE, Grant DF. Distribution and expression of soluble epoxide hydrolase in human brain. J Histochem Cytochem 2008; 56: 551–559.
60 Rigaud D, Tallonneau I, Verges B. Hypercholesterolaemia in anorexia nervosa: frequency and changes during refeeding. Diabetes Metab 2009; 35: 57–63.
61 Weinbrenner T, Zuger M, Jacoby GE, Herpertz S, Liedtke R, Sudhop T et al. Lipoprotein metabolism in patients with anorexia nervosa: a case-control study investigating the mechanisms leading to hypercholesterolaemia. Br J Nutr 2004; 91: 959–969.
62 Sato K, Emi M, Ezura Y, Fujita Y, Takada D, Ishigami T et al. Soluble epoxide hydrolase variant (Glu287Arg) modifies plasma total cholesterol and triglyceride phenotype in familial hypercholesterolemia: intrafamilial association study in an eight-generation hyperlipidemic kindred. J Hum Genet 2004; 49: 29–34.
63 Imig JD. Epoxides and soluble epoxide hydrolase in cardiovascular physiology. Physiol Rev 2012; 92: 101–130.
64 Mertes I, Fleischmann R, Glatt HR, Oesch F. Interindividual variations in the activities of cytosolic and microsomal epoxide hydrolase in human liver. Carcinogenesis 1985; 6: 219–223.
65 Iyer A, Kauter K, Alam MA, Hwang SH, Morisseau C, Hammock BD et al. Pharmacological inhibition of soluble epoxide hydrolase ameliorates diet-induced metabolic syndrome in rats. Exp Diabetes Res 2012; 2012: 758614.
66 De Taeye BM, Morisseau C, Coyle J, Covington JW, Luria A, Yang J et al. Expression and regulation of soluble epoxide hydrolase in adipose tissue. Obesity (Silver Spring) Mar 18: 489–498.

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