SELF-ARCHIVING VERSION

Anorexia Nervosa Genetics Initiative (ANGI) Identifies Eight Loci Associated with Anorexia Nervosa

Characterized primarily by extremely low BMI, anorexia nervosa (AN) is a complex, serious, and commonly misunderstood illness[1]. AN predominantly affects women, carries high risks of mortality from cachexia and suicide[2,3], and treatment outcomes remain unacceptably poor[4]. No medications exist that are effective and improved treatments will likely depend on deeper understanding of the core biology of this illness[5]. Genetic factors clearly play an etiological role with twin-based heritability estimates of 50-60%[6]. Combining samples from the Anorexia Nervosa Genetics Initiative (ANGI)[7,8] and the Eating Disorders Working Group of the Psychiatric Genomics Consortium (PGC-ED), we conducted a genome-wide association study (GWAS) meta-analysis of clinical, population, and volunteer cohorts and identified 8 independent genome-wide significant loci in 16,991 AN cases and 56,059 controls. Our analyses reveal that the genetic architecture of AN mirrors its clinical features and comorbidities, with high genetic correlations with obsessive-compulsive disorder (OCD), major depressive disorder (MDD), and anxiety disorders. Notably, the etiology of AN had the strongest metabolic and anthropometric genetic components of any psychiatric disorder yet examined. Specifically, we observed strong negative genetic correlations between AN and fasting insulin levels, body fat percentage, and BMI as well as a strong positive genetic correlation between AN and high-density lipoprotein cholesterol. These results support a broadened reconceptualization of AN etiology, underscoring that AN should be considered as both a psychiatric and metabolic disorder. Deeper understanding of the metabolic component is a critical next step and attention to both components may be necessary to improve treatment efficacy.

AN affects 0.9-4% of women and 0.3% of men with high medical and psychiatric morbidity[9,10]. Mortality rates associated with AN are elevated fivefold and are the highest of any psychiatric disorder[11]. There are few effective psychotherapeutic interventions, especially for adults, and pharmacological interventions—typically directed toward management of secondary symptoms—are minimally effective for the core features of AN[5]. Using genetics to elucidate the biological basis of AN could spur the development of effective treatments and reduce mortality.

Genetic factors play a role in AN[6]. The 2017 PGC-ED GWAS of 3,495 cases and 10,982 controls, estimated the common genetic variant-based heritability as ~20%, identified the first genome-wide significant locus, and reported significant genetic correlations (r_g) between AN and psychiatric and metabolic phenotypes[12]. Genetic correlations are not affected by reverse causation, although they may be mediated[13], and these results hinted at a role for metabolic etiological factors. In this study, we quadrupled the number of AN cases by combining the Anorexia Nervosa Genetics Initiative (ANGI)[7,8], the Genetic Consortium for Anorexia Nervosa/Wellcome Trust Case Control Consortium 3[14], and the UK Biobank[15]. Our goals were to identify convincing genotype-phenotype associations and elucidate the biology underlying AN.

Our GWAS meta-analysis included 33 datasets comprising 16,991 cases and 56,059 controls of European ancestry from 17 countries (**Supplementary Tables 1-4**). As typically seen in large GWAS of complex traits, we observed inflation of the test statistics (λ = 1.22, **Supplementary Figure 1**) consistent with polygenicity but no evidence of significant population stratification (See **Supplementary Results**). Meta-analysis results were completed for the autosomes and the X chromosome and were clumped to define eight independent genetic loci exceeding genome-wide significance ($P < 5 \times 10^{-8}$; **Figure 1; Table 1**) with consistent effects across cohorts. Seven of the loci have been associated in GWAS of other phenotypes, including autoimmune disorders, educational attainment, age at menarche, HDL cholesterol, and BMI (**Supplementary Table 5)**. **Supplementary Tables 6-8** describe annotations of genes within each locus. **Supplementary Table 9** reports a look-up restricted to the single-gene loci.

The top single nucleotide polymorphism (SNP) was rs9821797 ($P = 4.53 \times 10^{-15}$; odds ratio (OR) = 1.17; 95% confidence interval (CI): 1.13-1.22) in a locus on chromosome 3 with reported associations including autoimmune diseases, educational attainment, and HDL cholesterol. The locus contains >50 genes (see **Supplementary Table 6** and **Supplementary Table 8**). The second most significant SNP (rs6589488; chr 11; $P = 9.57 \times 10^{-11}$; OR = 1.13; 95% CI: 1.09-1.18) has been associated with BMI, and maps to *CADM1* (cell adhesion molecule 1), a member of the immunoglobulin superfamily with hypothesized roles in behavioral phenotypes[16]. The third locus (rs2287348; chr 2; $P = 4.68 \times 10^{-9}$; OR = 1.11; 95% CI: 1.07-1.15) maps on to seven genes and has previously been associated with BMI. The fourth locus (rs12415589; chr 10; $P = 1.80 \times 10^{-8}$; OR = 1.08; 95% CI: 1.05-1.11) maps on to *MGMT* (O^6-methylguanine–DNA methyltransferase) which is involved in DNA repair[17]. The fifth locus (rs370838138; chr 5; $P = 1.84 \times 10^{-8}$; OR = 1.08, 95% CI: 1.05-1.11) is intergenic and not associated with any other trait. The sixth locus (rs9874207; chr 3; $P = 2.44 \times 10^{-8}$; OR = 1.08; 95% CI: 1.05-1.12) maps on to *FOXP1* (forkhead box P1), rare mutations which cause intellectual disability (OMIM #613670) and has been associated with esophageal adenocarcinoma. The seventh locus (rs13125932; chr 4; $P = 4.64 \times 10^{-8}$; OR = 1.08; 95% CI: 1.05-1.10) is intergenic. The eighth locus (rs34816338; chr 1; $P = 4.98 \times 10^{-8}$; OR = 1.08; 95% CI: 1.05-1.12) was not previously associated. For detailed references see **Supplementary Results**. The previous locus reported in Duncan *et al*.[12] on 12q13.2 did not reach genome-wide significance ($P = 8.85 \times 10^{-6}$); however, between-cohort heterogeneity was apparent ($I^2 = 53.74$; **Supplementary Results** and **Supplementary Figure 2**).

Similar to the work by Won *et al*.[18], we used Hi-C data generated from human adult brain to identify genes implicated by three-dimensional functional interactomics (**Methods** and **Supplementary Figure 3 a-h**). These Hi-C data (N = 3, anterior temporal cortex) contain 509K high-confidence chromatin interactions (P Giusti-Rodríguez and PF Sullivan, personal communication). First, three multigenic loci were highly complex with many brain-expressed genes and a dense pattern of chromatin interactions precluding the identification of any single gene (locus 1, chr3:47.5-51.6 Mb, $P = 4.53 \times 10^{-15}$; locus 3, chr2:53.8-54.3 Mb, $P = 4.7 \times 10^{-9}$; locus 8, chr1:192.8-193.4, $P = 4.97 \times 10^{-8}$). Second, functional interactomic data from brain confirmed that the two single-gene associations were to the genes in which the associations were located. For locus 2 (chr11:114.9-115.4 Mb, $P = 9.57 \times 10^{-11}$), the vast majority of chromatin interactions were from the association region to *CADM1* (936 of 1051 interactions). For locus 4 (chr10:131.2-131.4 Mb, $P = 1.80 \times 10^{-8}$), an intragenic association in *MGMT*, the only chromatin interactions were to the same gene. Third, for the intergenic associations, locus 5 (chr5:24.9-

25.37 Mb, $P = 1.83 \times 10^{-8}$) had chromatin interactions with *CDH10* (cadherin 10) suggesting a functional connection, and all 79 chromatin interactions for locus 6 (chr3:70.6-71.0, $P = 2.43 \times 10^{-8}$) were to *FOXP1*. Locus 7 (chr4:58.0-58.2, $P = 4.64 \times 10^{-8}$) had no clear connection. These data suggest that *CADM1*, *MGMT*, *CDH10*, and *FOXP1* may play a role in the etiology of AN.

Conditional and joint analysis (COJO) of our genome-wide significant SNPs confirm their independence[19] (**Supplementary Table 10 and Methods**). **Supplementary Table 11** shows the results of multi-trait analysis (mtCOJO)[20] conditioning our genome-wide significant SNPs on variants that were associated in GWASs of BMI, HDL, type 2 diabetes, schizophrenia, neuroticism, and education years. Five loci appear to be independent. However, conditioning on neuroticism revealed one additional genome-wide significant hit on chromosome 1 (rs10747478, chr 1; $P = 3.87 \times 10^{-8}$; OR = 1.08; 95% CI: 1.05-1.11). The loci on chromosomes 1, 4, and 10 may not be independent of the associations identified by the GWAS that were used for the conditioning.

The liability-scale SNP heritability (SNP-h^2) of these results using linkage disequilibrium score regression (LDSC)[21] was 13% (s.e. = 1%) assuming a population prevalence of 0.9%, supporting the polygenic nature of AN. Polygenic risk score (PRS) analyses were conducted using a leave-one-dataset-out approach and indicated that the PRS captures ~1.7% ($P = 1.62 \times 10^{-13}$) of the phenotypic variance on the liability scale. We did not observe differences in polygenic architecture between males and females with AN (male: $N = 447$ cases $N = 20{,}347$ controls) or between binge eating and non-binge eating subtypes of AN (**Methods**, **Supplementary Results**, **Supplementary Figure 4-5**); however, these are tentative results and await larger samples and increased power before conclusions can be drawn.

We then tested SNP-based genetic correlations (SNP-r_g) with a wide range of external traits and disorders using bivariate LDSC[21]. Significant SNP-r_g surpassing Bonferroni correction fell into five categories: psychiatric and personality traits; physical activity; anthropometric traits; metabolic traits; and educational attainment (**Figure 2, Supplementary Tables 12-13**). Positive SNP-r_gs were observed with OCD (SNP-$r_g \pm$ s.e. = 0.45 ± 0.07; $P = 1.12 \times 10^{-9}$), MDD (0.25 ± 0.06; $P = 8.77 \times 10^{-5}$), anxiety disorders (0.25 ± 0.04; $P = 8.39 \times 10^{-9}$), schizophrenia (0.24 ± 0.02; $P = 1.67 \times 10^{-22}$), and bipolar disorder (0.19 ± 0.05; $P = 8.65 \times 10^{-5}$). This pattern reflects observed comorbidities in clinical and epidemiological studies[22,23]. The positive SNP-r_g with physical activity (0.17 ± 0.04; $P = 4.78 \times 10^{-5}$) is a new finding that encourages further exploration of the perplexing and difficult to treat symptom of pathologically elevated activity in AN[24]. We observed significant positive SNP-r_g with educational attainment (0.21 ± 0.04; $P = 7.73 \times 10^{-9}$) and related constructs, but no significant r_g with IQ. Expanding previous observations[12], we observed a palette of metabolic and anthropometric genetic correlations with AN far more pronounced than that seen in other psychiatric disorders. We observed significant negative SNP-r_g with fat mass (-0.33 ± 0.03; $P = 2.74 \times 10^{-24}$), fat free mass (-0.12 ± 0.03; $P = 4.09 \times 10^{-5}$), BMI (-0.31 ± 0.03; $P = 5.38 \times 10^{-23}$), obesity (-0.22 ± 0.04; $P = 1.77 \times 10^{-10}$), fasting insulin (-0.25 ± 0.06; $P = 3.42 \times 10^{-5}$), insulin resistance (-0.29 ± 0.07; $P = 6.74 \times 10^{-5}$), type 2 diabetes (-0.22 ± 0.05; $P = 2.75 \times 10^{-5}$), and leptin (-0.26 ± 0.07; $P = 1.00 \times 10^{-4}$) along with a significant positive SNP-r_g with HDL cholesterol (0.22 ± 0.04; $P = 4.34 \times 10^{-8}$).

With an increased number of GWAS loci for AN, systems biology analyses of our results revealed interesting albeit preliminary results. Firstly, partitioned heritability analysis showed, as with other GWAS[25], enrichment of SNP-h^2 in conserved regions (fold enrichment = 25.0, s.e. 3.3, $P = 1.12 \times 10^{-11}$; **Supplementary Figure 6**)[26]. Cell type group-specific annotations revealed that the overall SNP-h^2 is significantly enriched for CNS tissue (**Supplementary Figure 7**). Second, we evaluated whether genes associated with AN were enriched in specific pathways, tissues or cell types (**Methods**) using MAGMA (**Supplementary Table 14**)[27]. One biological pathway was significant, but its p-value was very close to the significance threshold (**Supplementary Table 15**): GO:positive_regulation_of_embryonic_development (Gene Ontology, 32 genes, $P = 6.31 \times 10^{-6}$). Genes associated with AN were enriched for expression specificity in the majority of brain tissues in GTEx, particularly the cerebellum and cerebellar hemispheres (**Supplementary Figure 8**), which have the highest proportion of neurons[28]. Similarly, among specific cell types, significant enrichment was found for pyramidal neurons from the CA1 region of the hippocampus (**Supplementary Figure 9**). Twenty-nine genes were predicted to be tissue-specific differentially expressed in 29 GTEx tissues (**Supplementary Table 16**) with *MGMT* on chromosome 10 being predicted to be downregulated in the caudate. We interpret these results cautiously, but they represent the first indications of the specific pathways, tissues, and cell types that mediate the genetic risk for AN). with *MGMT* on chromosome 10 being predicted to be downregulated in the caudate. We interpret these results cautiously, but they represent the first indications of the specific pathways, tissues, and cell types that mediate the genetic risk for AN.

In conclusion, we report multiple genetic loci with promising clinical and functional analyses and enrichments. The most striking finding is a strengthened pattern of both psychiatric and metabolic/anthropometric genetic correlations, which inspire new lines of inquiry into the metabolic and anthropometric components of AN etiology and disorder maintenance.

Low BMI and associated traits have traditionally been viewed as symptoms of AN. This perspective has failed to yield any interventions that reliably lead to sustained weight gain and recovery. Our results suggest that metabolic factors may play an etiological role in AN or that many of the same genes may influence both AN and metabolic traits. This was hypothesized decades ago[29], but is now supported by genetic data. Clinically, AN is nearly universally characterized by exceptional difficulty in maintaining a healthy BMI (even after therapeutic renourishment) in the face of an overwhelming biological and psychological pull toward low weight. Addressing metabolic rather than purely psychological drivers of this process may open new avenues for treatment and inspire novel biological approaches to understanding and treating this frequently lethal illness. Genomic discovery in AN is underway and will accelerate with increasing sample size.

URLs. LDhub, http://ldsc.broadinstitute.org**;** LDSC, https://github.com/bulik/ldsc; MAGMA, http://ctg.cncr.nl/software/magma; PRSice, https://github.com/choishingwan/PRSice/wiki; Ricopili, https://github.com/Nealelab/ricopili/tree/master/rp_bin; UK Biobank, http://www.ukbiobank.ac.uk

METHODS

Methods are available in the online version of the paper.

ACKNOWLEDGMENTS

Grant support for the Anorexia Nervosa Genetics Initiative, the Eating Disorder Workgroup of the Psychiatric Genomics Consortium, and its component groups are shown in **Supplementary Table 17**.

We thank all study volunteers, study coordinators, and research staff who enabled this study. ANGI: The Anorexia Nervosa Genetics Initiative (ANGI) was an initiative of the Klarman Family Foundation. Additional support was offered by the National Institute of Mental Health. PGC: We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of individuals who have shared their life experiences with PGC investigators and the contributing studies. We are grateful to the Children's Hospital of Philadelphia (CHoP), the Price Foundation Collaborative Group, the Genetic Consortium for Anorexia Nervosa (GCAN), the Wellcome Trust Case Control Consortium-3 (WTCCC-3), the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), the QSkin Sun and Health Study, Riksät (Swedish National Quality Register for Eating Disorders), the Stockholm Center for Eating Disorders (SCÄ), LifeGene, the United Kingdom Biobank, and all PGC-ED members for their support in providing individual samples used in this study. We thank SURFsara (http://www.surf.nl) for support in using the Lisa Compute Cluster. We thank Max Lam, Institute of Mental Health, Singapore, for Ricopili consultation. This study also represents independent research partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

AUTHOR CONTRIBUTIONS

C.B. and P.F.S. conceived and designed the study. L.T., C.B., & G.B., performed overall study coordination. C.B., P.F.S., N.M., M.L., P.M., P.L., A.B., C.N. were PIs of ANGI. H.W., Z.Y., J.C., C.H., J.B., H.G., S.Y., V.L., & M.M. performed the statistical analyses. S.R. provided statistical consultation. A.H. assisted with data interpretation. C.B., L.T., J.J., M.K., G.M., T.W., A. B., P.L., & C.N. collected and managed the ANGI samples at sites. C.B., M.L., N.M., and P.M. are ANGI site PIs. D.W., C.O., T.W., W.K., A.K., B.W., J.M., C.J., H.B., S.C., J.H., J.B., J.P., N.P., A.F., & G.B. contributed to data acquisition. C.B., G.B., & P.S. supervised the study. H.W., C.B., Z.Y., C.H., G.B., J.C., H.G., S.Y., & J.B. wrote the manuscript. Members of the Psychiatric Genomics Consortium Eating Disorder workgroup (list initials in alphabetical order) made individual data from subjects available for inclusion

COMPETING FINANCIAL INTERESTS

CM Bulik reports: Shire (grant recipient, Scientific Advisory Board member) and Pearson and Walker (author, royalty recipient). G Breen has received grant funding from and served as a consultant to Eli Lilly and has received honoraria from Illumina. PF Sullivan reports: Lundbeck (advisory committee), and Roche (grant recipient, speaker reimbursement). HJ Watson, Z Yilmaz, C Hübel, HA Gaspar, JRI Coleman have nothing to declare.

References

1. Schaumberg, K. *et al.* The science behind the Academy for Eating Disorders' nine truths about eating disorders. *Eur. Eat. Disord. Rev.* **25**, 432-450 (2017).
2. Fichter, M.M. & Quadflieg, N. Mortality in eating disorders - results of a large prospective clinical longitudinal study. *Int. J. Eat. Disord.* **49**, 391-401 (2016).
3. Yao, S. *et al.* Familial liability for eating disorders and suicide attempts: evidence from a population registry in Sweden. *JAMA Psychiatry* **73**, 284-291 (2016).
4. Watson, H. & Bulik, C. Update on the treatment of anorexia nervosa: review of clinical trials, practice guidelines and emerging interventions. *Psychol. Med.* **43**, 2477-2500 (2013).
5. Treasure, J. *et al.* Anorexia nervosa. *Nat. Rev. Dis. Primers* **1**, 15074 (2015).
6. Yilmaz, Z., Hardaway, J.A. & Bulik, C.M. Genetics and epigenetics of eating disorders. *Adv. Genomics Genet.* **5**, 131-150 (2015).
7. Kirk, K.M. *et al.* The Anorexia Nervosa Genetics Initiative: study description and sample characteristics of the Australian and New Zealand arm. *Aust. N. Z. J. Psychiatry* **51**, 583-594 (2017).
8. Thornton, L.M. *et al.* The Anorexia Nervosa Genetics Initiative: overview and methods. *bioRxiv* (2017).
9. Keski-Rahkonen, A. & Mustelin, L. Epidemiology of eating disorders in Europe: prevalence, incidence, comorbidity, course, consequences, and risk factors. *Curr. Opin. Psychiatry* **29**, 340-345 (2016).
10. Hudson, J.I., Hiripi, E., Pope, H.G. & Kessler, R.C. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol. Psychiatry* **61**, 348-358 (2007).
11. Arcelus, J., Mitchell, A.J., Wales, J. & Nielsen, S. Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. *Arch. Gen. Psychiatry* **68**, 724-731 (2011).
12. Duncan, L. *et al.* Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **173**, 850-858 (2017).
13. Pickrell, J.K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709-717 (2016).
14. Boraska, V. *et al.* A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085 (2014).
15. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
16. Patak, J., Zhang-James, Y. & Faraone, S.V. Endosomal system genetics and autism spectrum disorders: a literature review. *Neurosci. Biobehav. Rev.* **65**, 95-112 (2016).
17. Hegi, M.E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Eng. J. Med.* **352**, 997-1003 (2005).
18. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).
19. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369-375 (2012).

20. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
21. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291-295 (2015).
22. Cederlöf, M. *et al.* Etiological overlap between obsessive-compulsive disorder and anorexia nervosa: a longitudinal cohort, multigenerational family and twin study. *World Psychiatry* **14**, 333-338 (2015).
23. Kaye, W.H., Bulik, C.M., Thornton, L., Barbarich, N. & Masters, K. Comorbidity of anxiety disorders with anorexia and bulimia nervosa. *Am. J. Psychiatry* **161**, 2215-2221 (2004).
24. Dalle Grave, R., Calugi, S. & Marchesini, G. Compulsive exercise to control shape or weight in eating disorders: prevalence, associated features, and treatment outcome. *Compr. Psychiatry* **49**, 346-352 (2008).
25. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228-1235 (2015).
26. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476 (2011).
27. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
28. Azevedo, F.A.C. *et al.* Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513**, 532-541 (2009).
29. Hebebrand, J. & Remschmidt, H. Anorexia nervosa viewed as an extreme weight condition: genetic implications. *Hum Genet* **95**, 1-11 (1995).