ARTICLE

Whole-exome sequencing reveals common and rare variants in immunologic and neurological genes implicated in achalasia

Quanlin Li,[1,5] Weifeng Chen,[1,5] Cheng Wang,[2,3,4,5] Zuqiang Liu,[1] Yayun Gu,[2,3] Xiaoyue Xu,[1] Jiaxing Xu,[1] Tao Jiang,[2,3] Meidong Xu,[1] Yifeng Wang,[2,3] Congcong Chen,[2,3] Yunshi Zhong,[1] Yiqun Zhang,[1] Liqing Yao,[1] Guangfu Jin,[2,3] Zhibin Hu,[2,3,*] and Pinghong Zhou[1,*]

Summary

Idiopathic achalasia (IA) is a severe motility disorder characterized by neuronal degeneration in the myenteric plexus, but the etiology remains largely unknown. We performed whole-exome sequencing (WES) in 100 IA-affected individuals and 313 non-IA control subjects and validated the results in 230 IA-affected individuals and 1,760 non-IA control subjects. Common missense variants rs1705003 (*CUTA*, GenBank: NC_000006.11:g.33385953A>G) and rs1126511 (*HLA-DPB1*, GenBank: NC_000006.11:g.33048466G>T) at 6p21.32 were reproducibly associated with increased risk of IA (rs1126511: OR = 1.83, $p = 2.34 \times 10^{-9}$; rs1705003: OR = 2.37, $p = 3.21 \times 10^{-7}$), meeting exome-wide significance. Both variants can affect the expression of their target genes at the transcript level. An array-based association analysis in 280 affected individuals and 1,121 control subjects determined the same signal at 6p21.32. Further conditional analyses supported that the two missense variants identified in WES-based association study were potential causal variants of IA. For rare variants, the top genes identified by gene-based analysis were significantly enriched in nerve and muscle phenotypic genes in the mouse. Moreover, the functional rare variants in these genes tended to cooccur in IA-affected individuals. In an independent cohort, we successfully validated three rare variants (*CREB5*, GenBank: NC_000007.13:g.28848865G>T; *ESYT3*, GenBank: NC_000003.11:g.138183253C>T; and *LPIN1*, GenBank: NC_000002.11:g.11925128A>G) which heightens the risk of developing IA. Our study identified and validated two common variants and three rare variants associated with IA in immunologic and neurological genes, providing new insight into the etiology of IA.

Introduction

Idiopathic achalasia (IA [MIM: 200400]) is a motility disorder characterized by esophageal aperistalsis and defective relaxation of the lower esophageal sphincter (LES). Although it has an annual incidence of only approximately 1/100,000, the disease is chronic, progressive, and could cause a substantially impaired ability to eat and lessened quality of life.[1,2] The impaired flow and stasis of ingested food can result in typical symptoms of dysphagia, regurgitation, retrosternal pain, and extreme weight loss, as well as an elevated risk of esophageal carcinoma (MIM: 133239),[2] and they greatly affect the health-related quality of life, work productivity, and functional status of affected individuals.

The histopathology of IA involves the degeneration of inhibitory nitrergic neurons in the myenteric plexus.[3] Although the specific pathogenesis of achalasia remains largely unclear, the hypothesis of autoimmune-mediated ganglionitis triggered by potential virus infections has been proposed to underlie the loss of myenteric neurons, particularly in genetically susceptible individuals.[2]

Familial achalasia and genetic syndrome accompanied by achalasia also supported that genetic factors play an important role in the pathogenesis of achalasia.[4,5] Subsequent genetic studies determined a number of common loci associated with IA.[6–8] Recently, a genome-wide association study (GWAS) in Europeans identified multiple common variants in the HLA-DQ region that confer susceptibility to IA.[9] Rare variants, however, were seldom investigated in the development of IA.

Almost all these studies were conducted in individuals with European ancestry. However, studies have shown genetic heterogeneity in IA susceptibility across populations.[6,10] It is biologically conceivable that the same susceptibility variant for IA may also be implicated in the Chinese population. However, it is also possible that variants identified in populations of European descent might not be applicable to Chinese populations because of underlying genetic heterogeneity.

Here, we performed a sequencing-based exome-wide association study of IA in a Chinese Han population and identified and validated that two independent common variants at 6p21.32 and three rare and functional variants (*CREB5* [MIM: 618262], GenBank: NC_000007.13:g.28848865G>T; *ESYT3* [MIM: 616692], GenBank: NC_000003.11:g.138183253C>T; and *LPIN1* [MIM: 605518], GenBank: NC_000002.11:g.11925128A>G) contribute to the

[1]Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China; [2]Department of Epidemiology and Biostatistics, Center of Global Health, School of Public Health, Nanjing Medical University, Nanjing 211166, China; [3]State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 211166, China; [4]Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211116, China
[5]These authors contributed equally
*Correspondence: zhibin_hu@njmu.edu.cn (Z.H.), zhou.pinghong@zs-hospital.sh.cn (P.Z.)
https://doi.org/10.1016/j.ajhg.2021.06.004.

risk of IA. Interestingly, those rare variants tended to locate in the genes affecting nerve and muscle function and to cooccur in IA-affected individuals.

Material and methods

Sample collection

Our study included 330 affected individuals with idiopathic achalasia undergoing peroral endoscopic myotomy (POEM) at the Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Shanghai, China. Patients with achalasia were basically diagnosed by high-resolution manometry, assisted by history and barium esophagogram. In cases where secondary achalasia could not be excluded, esophagogastroscopy and/or computed tomography or nuclear magnetic resonance (CT/NMR) imaging was conducted. Clinical characteristics were prospectively collected by medical history questionnaires. The 2,073 control subjects were selected from those receiving routine physical examinations in local hospitals or those participating in a community screening of noncommunicable diseases. The summary description of the samples used in this study is provided in Table S1.

Blood samples from patients and controls for DNA analysis were collected after acquiring written informed consent. The 10 mL samples were stored at −80°C immediately after collection. The study was reviewed and approved by the local research ethics committee.

Whole-exome sequencing

DNA was extracted from blood samples using the QIAamp DNA Blood & Tissue Kit (69506). Library construction and whole-exome capture of genomic DNA were performed using the SureSelectXT reagent kit (Agilent) and SureSelectXT Human All Exon V5 (Agilent) on an Agilent Bravo Automated Liquid Handling Platform (Agilent). The captured DNA was sequenced on an Illumina HiSeq 2500 sequencing platform, with 150-bp paired-end sequencing.

Sequencing alignment and variant detection

The FastQC package was used to assess the quality-score distribution of the sequencing reads. Read sequences were mapped to the human reference genome (GRCh37) using the Burrows-Wheeler Aligner (BWA-MEM v.0.7.15-r1140)[11] with the default parameters, and duplicates were marked and discarded using Picard (v.1.70). After alignment by BWA, the reads were subjected to recalibration using the Genome Analysis Toolkit (GATK v.3.8).[12] Haplotype calling was performed using HaplotypeCaller in GATK v.3.8 in GVCF mode according to the best practice and all GVCF were merged and joint genotyped with GenotypeGVCFs in GATK v.3.8 to produce a combined VCF file for further analysis. We applied xHLA with default parameters to conduct Human Leukocyte Antigen (HLA) typing on whole-exome sequencing (WES) data after re-aligning the reads to the human reference genome (GRCh38).[13]

Quality control

Before variant calling, we applied sequencing quality control on 100 IA-affected individuals and 323 normal control subjects (Figure S1). Five subjects were excluded from further analysis if (1) the average sequence depth $< 60\times$ or (2) they were contaminated, as defined by Freemix > 0.05. VerifyBamID v.1.0.5 was used for contamination estimates. After variant calling, we used SnpEff[14] for gene annotation and vcfanno[15] for variant annotation. Gene information from the Gencode V19 was used as a reference for gene annotation and the population frequency in the Exome Aggregation Consortium (ExAC) v.0.3.1 was used for variant annotation.

Then, we applied standard variant-level and individual-level quality controls. We first applied variant quality score recalibration (VQSR) on detected variants using GATK v.3.8 according to the best practice. Then, we excluded variants for further analysis if (1) they were not in the coding region; (2) they were in the sex chromosome; (3) they failed in VQSR evaluation; (4) they were located in the segmental duplication region marked by the UCSC browser; (5) they had calling rates $< 95\%$; or (6) the genotype distributions of the SNPs deviated from those expected by Hardy-Weinberg equilibrium. After that, we excluded samples with calling rates < 0.95. After the implementation of all quality controls mentioned above, we included 100 IA-affected individuals and 313 control subjects in the final analysis. After variant annotation, a total of 192,329 coding variants were identified of high quality and were further classified according to their minor allele frequency (MAF) into 33,687 common variants ($MAF_{case} > 0.01$ or $MAF_{control} > 0.01$) and 158,642 rare variants ($MAF_{case} \leq 0.01$ and $MAF_{control} \leq 0.01$).

We detected population outliers and stratification using a method based on principal component analysis,[16] which indicated that the affected individuals and control subjects were genetically matched for all sequenced samples (Figure S2A). The quantile-quantile plot revealed a good match between the distributions of the observed p values and those expected by chance (Figure S2B). The small genomic control inflation factor (λ) of 0.92 also indicated a low possibility of false-positive associations resulting from population stratification.

Common SNP selection for validation

Exome-wide association analysis on common SNPs was performed using an additive model in a logistic regression (1 degree of freedom) analysis by PLINK2. After that, common SNPs were selected for further validation based on the following criteria: (1) SNPs had $p \leq 1.0 \times 10^{-4}$ and (2) only the SNP with the lowest p value was selected when multiple SNPs were observed that were in strong LD ($r^2 \geq 0.8$). As a result, five SNPs met criterion 1, and among them, two SNPs met criterion 2; the other three SNPs were in high LD with rs1126504 (GenBank: NC_000006.11:g.33048457C>G) (Figure S3).

Gene expression and expression quantitative trait loci (eQTL) analyses

Fully processed, filtered, and normalized gene expression matrices for each tissue and official matrix of covariates used in eQTL analysis were downloaded from the Genotype-Tissue Expression (GTEx) data portal. The corresponding genotype information was downloaded from dbGaP with accession number phs000424.v6.p1. A linear regression model was used in the eQTL analysis to evaluate the association between a SNP and the gene expression of 500 Kb up- and downstream of the SNP, adjusting for the covariates involved in the standard GTEx analysis (i.e., top 3 genotyping principal components, PEER factors determined by sample size, genotyping platform, and sex).

Gene-based analysis and cooccurrence analysis

A total of 158,642 rare variants that passed QC were included in the following gene-based analysis. We conducted gene-based

analysis using the combined tests of burden test and SNP-set (sequence) kernel association test (SKAT-O) as implemented in the R-package SKAT. Because IA is a rare disease, we also applied functional variant clustering analysis with the hypothesis that the functional IA-specific variants accumulate in the IA-related genes. We first defined IA-specific SNPs if they satisfied (1) ExAC $MAF_{All} < 0.05$; (2) ExAC $MAF_{Asian} < 0.05$; and (3) $MAF_{control} = 0$. Then, an *in silico* prediction based on Combined Annotation Dependent Depletion (CADD, GRCh37-v1.4)[17] scores were annotated to all variants. For each gene, a combined score was calculated as the summation of SNP recurrence multiplied by the CADD score and a p value was computed by testing whether the observed combined score is significantly higher than the null distribution. We built a null distribution by randomly sampling the same numbers of SNP positions from other protein-coding genes. The top 200 genes in SKAT-O and functional variant clustering analysis were used for gene set enrichment analysis by the hypergeometric test. Mouse phenotype annotation information for enrichment analysis was downloaded from Mouse Genome Informatics (MGI).[18]

Genes with $p < 0.05$ in both the SKAT-O test and the functional variant clustering test were included for cooccurrence analysis. We involved functional (CADD score > 15), rare, and IA-specific SNPs and calculated the cooccurring score as the frequency of cooccurring pairs. Then, we built a null distribution by randomly sampling the same numbers of SNP positions from other protein-coding genes. A p value was computed by testing whether the observed cooccurring pair frequency was significantly higher than the null distribution.

Rare SNP selection for validation

We used the Fisher method to obtain a combined p value from the SKAT-O test and the functional variant clustering test. The Benjamini-Hochberg method for multiple hypothesis correction and q value cut-off of 0.20 was used. Then, we involved SNPs in the following validation if (1) they were functional (CADD score > 15); (2) they were rare variants; (3) they were IA-specific variants; and (4) they were recurrent (occurred in at least two IA-affected individuals).

Validation by Sanger sequencing

Approximately 300 bp DNA sequence fragments containing the SNPs for validation were amplified by PCR using the primers in Table S2. PCR products were purified and sequenced by Sanger sequencing at TSINGKE Biological Technology (Nanjing). Sequences were compared with NCBI to build 37 references sequence by DNAMAN and Chromas software. The technicians who performed PCR amplification experiments were blinded to the case-control status of the samples.

Array-based genome-wide association analysis

Given the small number of common variants detected by WES, we further applied genome-wide SNP arrays to genotype a qualified subset of our discovery and validation samples and conducted classical genome-wide association analyses to confirm whether there were other common non-coding variants strongly associated with risk of IA. A total of 282 affected individuals and 1,128 control subjects with enough DNA were genotyped using Illumina Infinium Asian Screening Array (v.1.0), which contains 659,184 markers. Standard quality control at both the sample level and variant level followed the approach described in the previous publications.[19] Briefly, samples with a call rate of <0.95 or gender discrepancy, or samples with extreme heterozygosity (6 SD from the mean) were excluded. Identity by descent (IBD) analysis was calculated using PLINK to detect cryptic relatedness. When a pair of samples had PI_HAT > 0.25, the member with a lower call rate was excluded. Variants with call rate ≥ 95% and < 98%, or poor Illumina intensity or clustering metrics, or deviation from the expected frequency as observed in the 1000 Genomes Project (the Phase III integrated variant set release, 504 East Asians) were selected for manual inspection. SNPs that failed cluster inspection were excluded. We further excluded variants with a call rate of < 0.95, Hardy-Weinberg equilibrium (HWE) p value of $< 1 \times 10^{-6}$, or minor allele frequency (MAF) of < 0.01. As a result, a total of 280 affected individuals and 1,121 control subjects with 476,650 SNPs were included in the subsequent analysis of the GWAS. The PCA and Q-Q plot of the genome-wide association test showed minimal evidence of population stratification (Figure S4).

Imputation was then performed at 6p21.32 region by SHAPEIT (v.2)[20] and IMPUTE (v.2)[21] with default parameters. The 1000 Genomes Project Phase III database was used as the reference. SNPs with imputation quality score INFO < 0.9 or MAF less than 0.01 were excluded from further analysis. The association analysis was done in the imputed dosage data using SNPTEST (v.2).[22] To detect independent association signals, we performed the stepwise conditional analyses at 6p21.32.

Statistical analyses

The baseline characteristics of groups are described in proportions for categorical variables (sex and achalasia stage) and with a mean (SD) for continuous variables (age and age of onset). For common variants, the meta-analysis was performed in a combined analysis with fixed-effect model and the p values were considered to be significant when below 0.05 after Bonferroni correction by multiplying the number of comparisons ($0.05/33{,}687 = 1.48 \times 10^{-6}$). For gene-based analysis, p values were adjusted by the Benjamini-Hochberg false discovery rate (FDR-BH). The eQTL analysis in GTEx data was processed according to its protocol using a linear regression model as described above. All statistical analysis was conducted using R 3.3.1 software if not otherwise specified.

Results

General description of our initial discovery stage based on whole-exome sequencing

After stringent quality control, we used WES data from 100 IA-affected individuals and 313 non-IA control subjects in the initial discovery stage (Table S1, Figure S1). After filtering and annotation of variants, a total of 192,329 coding variants were used for further analysis, including 33,687 common variants and 158,642 rare variants. Among these IA-affected individuals, we did not identify any variants previously reported in familial IAs or rare loss-of-function variants in genes with achalasia phenotype in a mouse model (Table S3).[23–33] Principal component analysis and quantile-quantile (Q-Q) plots of genome-wide association test statistics revealed no inflation; thus, substantial cryptic population substructure and differing genotypic variants between the affected

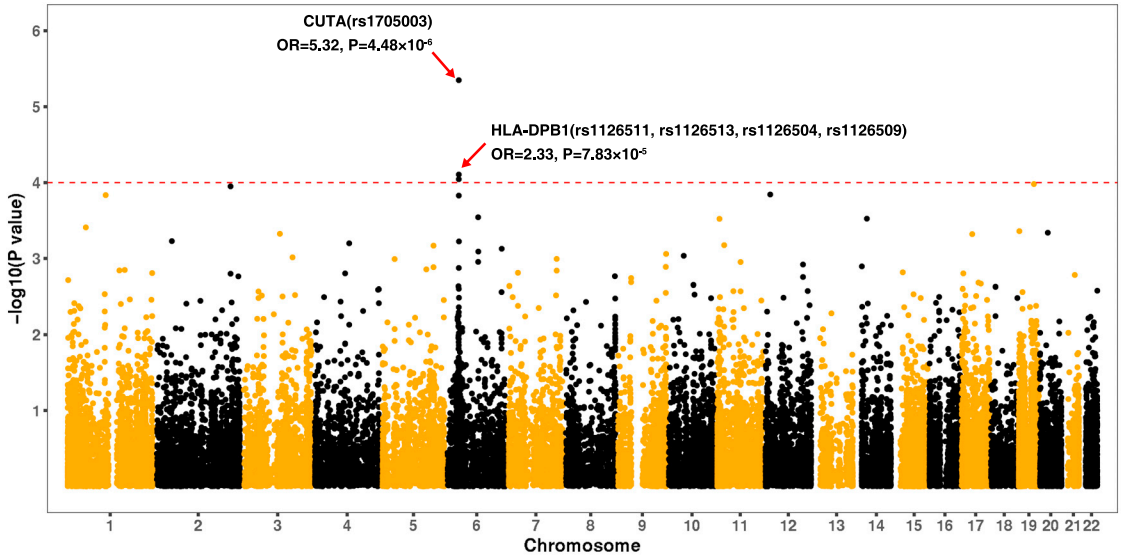

Figure 1. Manhattan plot of results from exome-wide association analysis in the discovery stage of 100 affected individuals and 313 control subjects
The red dashed line represents the significance threshold for validation ($p = 1 \times 10^{-4}$).

individuals and control subjects were unlikely in the discovery stages ($\lambda = 0.92$, Figure S2B).

Common variants at 6p21.32 were associated with IA

Among the common variants, five SNPs at the HLA region passed the criteria in our discovery stage ($p < 1 \times 10^{-4}$, see material and methods, Figure 1). We selected the variant rs1126511 (GenBank: NC_000006.11:g.33048466G>T) in *HLA-DPB1* (MIM: 142858) and the variant rs1705003 (GenBank: NC_000006.11:g.33385953A>G) in *CUTA* (MIM: 616953) for subsequent validation. Variant rs1705003 in *CUTA* is a missense variant in transcript ENST00000374500 and is also annotated in the promoter region of the other transcripts as it is located very close to the 5′ UTR (Figure S5). Because rs1126511 and rs1126513 are located in the same codon and are completely linked ($r^2 = 1$), the variant haplotype predicts a Gly to Leu change. We ignored the other three missense SNPs rs1126504 (GenBank: NC_000006.11:g.33048457C>G), rs1126513 (GenBank: NC_000006.11:g.33048467G>T), and rs1126509 (GenBank: NC_000006.11:g.33048461T>A) because of their high linkage disequilibrium ($r^2 > 0.8$) with rs1126511 (Figure S3).

In an independent validation with 230 IA-affected individuals and 1,760 non-IA control subjects, both variants were immediately validated and contributed to a consistent risk of IA ($OR_{rs1126511} = 1.71$, $p_{rs1126511} = 3.18 \times 10^{-6}$; $OR_{rs1705003} = 1.90$, $p_{rs1705003} = 7.47 \times 10^{-4}$, Table 1). In the combined analysis of the results of the discovery study and the validation study, both of the variant rs1126511 at the second exon of *HLA-DPB1* (ENST00000418931) and rs1705003 at the first exon of the *CUTA* transcripts (ENST00000374500) had strong association with risk of IAs ($OR_{rs1126511} = 1.83$, $p_{rs1126511} = 2.34 \times 10^{-9}$; $OR_{rs1705003} = 2.37$, $p_{rs1705003} = 3.21 \times 10^{-7}$, meeting exome-wide significance ($p < 0.05/33{,}687 = 1.48 \times 10^{-6}$).

Given the small number of common variants detected by WES, we further applied genome-wide SNP arrays to genotype a qualified subset of our discovery and validation samples and conducted classical genome-wide association analyses (280 IA-affected individuals and 1,121 control subjects, material and methods) to confirm whether there were other common variants strongly associated with risk of IA. Consistent with the previous Immunochip study in the European ancestry[9] and our association study based on WES, we identified a signal at 6p21.32 with suggestive genome-wide significance ($p < 1 \times 10^{-5}$, Figure S6). To further determine the independent signals at the locus, we performed imputation in the region based on the haplotypes from the 1000 Genomes Project Phase 3 database and conducted stepwise conditional analyses in the 6p21.32 region using both the variants identified in WES and the array with the p value $< 1 \times 10^{-4}$ (Figure S7 and Table S4). The top variant rs3097671 (GenBank: NC_000006.11:g.33047612G>C) identified by the array-based study was in moderate linkage disequilibrium with rs1126511 identified by the WES-based study in *HLA-DPB1* ($r^2 = 0.39$). After correction for rs3097671, the strongest independent signal was rs2772372 (GenBank: NC_000006.11:g.33427350T>C) identified by the WES-based study (Figure S7), which was in moderate linkage disequilibrium with rs1705003 in *CUTA* ($r^2 = 0.41$). Adjusting for rs3097671 and rs2772372, we found no additional independent signals in the region (Figure S7). The IA-risk SNP variant rs28688207 (GenBank: NC_000006.11:g.32628660T>C) reported in the European ancestry by Gockel et al.[9] showed a nominal significance (OR = 1.49, $p = 4.62 \times 10^{-3}$) with IA risk and was in low LD ($r^2 < 0.1$) with both the coding variants identified in this study (rs1126511 and rs1705003). After the correction for the reported rs28688207, the association of rs1126511 ($OR_{adj} =$

Table 1. The association results of common variants rs1126511 and rs1705003

				Discovery (WES)				Validation (SANGER Sequencing)					
Variant	Position	Gene	EA	EAF (subjects, n = 100)	EAF (controls, n = 313)	OR	p	EAF (subjects, n = 230)	EAF (controls, n = 1,760)	OR	p	OR (Meta)	p (Meta)
rs1126511	chr6:33048466	*HLA-DPB1*	T	0.305	0.165	2.33	7.83×10^{-5}	0.270	0.181	1.71	3.18×10^{-6}	1.83	2.34×10^{-9}
rs1705003	chr6:33385953	*CUTA*	G	0.115	0.027	5.32	4.48×10^{-6}	0.085	0.048	1.90	7.47×10^{-4}	2.37	3.21×10^{-7}

The discovery stage included 100 affected individuals and 313 control subjects and the validation stage included 230 affected individuals and 1,760 control subjects. EA, effect allele; EAF, effect allele frequency; OR, odds ratio.

1.75, $P_{adj} = 7.58 \times 10^{-7}$) and rs1705003 ($OR_{adj} = 2.23$, $P_{adj} = 1.93 \times 10^{-5}$) remained significant. Thus, we identified two independent signals at 6p21.32.

The common variants affect the protein sequence of *CUTA* and *HLA-DPB1* as well as their expression

Next, we evaluated the function of the variants on their host genes. The A-to-G base change of rs1705003 is at a conservative site and leads to a protein change from serine to proline. In addition to the amino acid change, the variant is also located in the regions with regulatory activity in neural progenitor cell lines (Figure S8). The eQTL analysis showed that the variant allele was associated with lower expression of *CUTA* in a total of 30 normal tissues, including brain and nerve tissues (Table S5). In the nerve tibial tissues, ENST00000374500 affected by rs1705003 was the fourth expressed transcript of *CUTA* (Figure S9). As rs1705003 was also located within 1 kb upstream of the transcription start site (TSS) of the primarily expressed transcript (ENST00000488034) of *CUTA*, we further investigate the association between rs1705003 and ENST00000488034. The eQTL analysis showed that the variant allele was associated with lower expression of ENST00000488034 in brain and nerve tissues (Table S6), suggesting the susceptibility mechanisms of rs1705003 could be complicated.

The variant rs1126511 affects ENST00000418931, which is the transcript of *HLA-DPB1* with the highest expression. It was also located in the promoter region of two transcripts (ENST00000428835 and ENST00000416804) in lymphocytes (GM12878) (Figure S10) and ENST00000428835 was the transcript with the second-highest expression in lymphocytes (Figure S11). We noticed that the T base was not only associated with lower expression ENST00000418931 in more than 20 tissues (Table S7) but also associated with higher expression of other transcripts of *HLA-DPB1* (ENST00000428835 and ENST00000416804) (Tables S8 and S9).

We also inferred the HLA loci according to the WES data and successfully determined the genotype of six HLA genes. We found that two *HLA-DPB1* loci (DPB1*21:01, DPB1*17:01) and one *HLA-DRB1* (MIM: 142857) locus (DRB1*07:01) contributed a risk effect on IA (Table S10).

To investigate the regulatory potential of variants moderately linked with rs3097671 and rs2772372 ($r^2 > 0.3$), we further performed functional annotation and found a number of variants that were located in the promoter or enhancer regions and had eQTL signals with nearby genes (Tables S11 and S12).

Rare variants in neurological and muscular genes were associated with IA and were prone to cooccur in cases of IA

For rare variants, we applied gene-based analysis in the whole-exome discovery study. As a linear combination of the burden and SKAT tests, SKAT-O tests whether a given gene has a high proportion of causal variants exerting effects in the same direction, or instead has many noncausal variants or variants exerting effects in opposite directions. Because IA is a rare disease, we mainly focused on rare variants dominant in IA-affected individuals in this study and conducted a functional variant clustering analysis, which hypothesized that IA-related genes carried more functional important and IA-specific variants in IA-affected individuals than the other genes. Interestingly, we observed that the genes with the nervous system and muscle phenotypes were consistently enriched in the top genes identified by the two methods (Figures 2A and 2B).

A total of 59 genes were considered as risk genes of IA because they were significantly associated with IA in SKAT-O analysis ($p_S < 0.05$) and showed a significantly higher number of IA-specific functional variants than expected ($p_F < 0.05$) (Table S13). In these risk genes, we identified 158 IA-specific rare variants with potential pathogenic effect ($Score_{CADD} > 15$). The variants involved 82 IA-affected individuals and were considered as potential IA risk variants. We then further investigated whether the IA risk variants in the risk genes contribute to the risk of IA cooperatively or independently. Interestingly, we found that the cooccurrence score of IA risk variants was significantly higher than that of randomly selected variants (Figure 2C), indicating that the IA risk variants may have a combined effect on the development of IA.

The replication of rare functional variants

Finally, we combined the p values of two gene-based analyses and then included six recurrent IA risk variants in seven

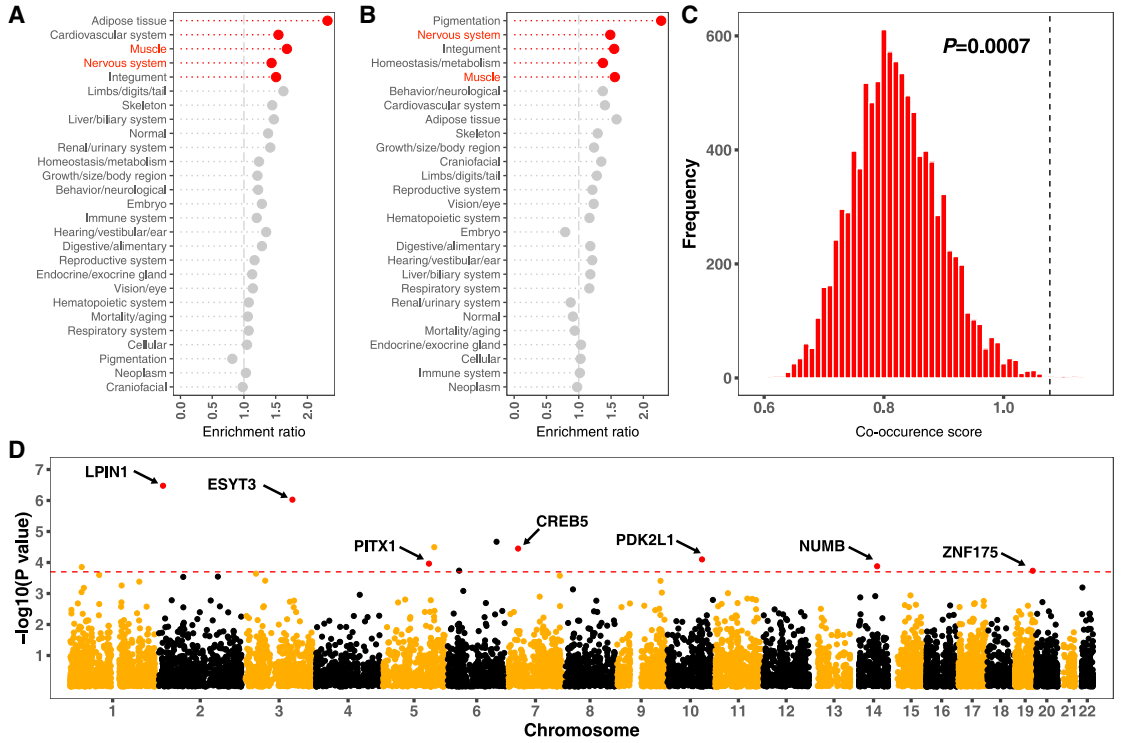

Figure 2. MGI gene set enrichment analysis on genes identified by exome-wide gene-based analysis and cooccurrence analysis
(A and B) The top 200 genes in SKAT-O and functional variant clustering analysis were included in the gene set enrichment analysis. Mouse phenotype annotation information from MGI was used as gene sets. The results were ranked by p value and the significant ones ($p < 0.05$) were marked as red.
(C) The number of cooccurring pairs of identified rare variants was significantly higher than that of the random variants. A null distribution was built by randomly sampling the same numbers of SNP positions from other protein-coding genes.
(D) Manhattan plot of combined p values from gene-based analysis. Seven genes were identified as they met the FDR threshold of 0.2 (red line).

genes with FDR-corrected combined p values < 0.2 (Figure 2D) into further validation including 230 affected individuals and 1,760 control subjects (Table 2). In this independent cohort, three rare variants (*CREB5* [MIM: 618262], GenBank: NC_000007.13:g.28848865G>T; *ESYT3* [MIM: 616692], GenBank: NC_000003.11:g.138183253C>T; and *LPIN1* [MIM: 605518], GenBank: NC_000002.11:g.11925128A>G) were successfully validated, as they occurred with significant enrichment in affected individuals compared with control subjects (Table 2). When the discovery and replicated cohort were combined, the most frequently found variant was a *CREB5* missense variant, which was present in 7/330 (2.1%) affected individuals and 7/2,073 (0.34%) control subjects (rs142716067, OR = 6.38, $p = 1.26 \times 10^{-3}$). A stop gained variant in *ESYT3* was also enriched in 5/330 (1.5%) IA-affected individuals versus 4/2,073 (0.19%) control subjects (rs770625121, OR = 7.95, $p = 3.72 \times 10^{-3}$). A missense variant in *LPIN1* was found in 5/330 (1.5%) IA-affected individuals and 5/2,073 (0.24%) control subjects (rs190743128, OR = 6.37, $p = 6.57 \times 10^{-3}$) (Table 2). In summary, 17/330 (5.2%) IA-affected individuals and 16/2,073 (0.77%) control subjects carried these three variants totally (OR = 6.97, $p = 2.15 \times 10^{-7}$). In contrast, all missense variants in other three genes (*PITX1* [MIM: 602149], *NUMB* [MIM: 603728], and *ZNF175* [MIM: 601139]) found in the affected individuals were also detected among control subjects without significant enrichment. Then, we evaluated the effects of missense variants by a deep neural network-based tool (DANN)[34] and a functional and conservation integration tool (FATHMM-MKL).[35] Both missense variants were predicted as deleterious with high probability (Table S14).

Discussion

Although IA is one of the most specific esophageal motor disorders, its etiology remains largely unknown. Previous studies supported the hypothesis that the neuronal degeneration of IA is caused by aberrant autoimmune responses triggered by specific environmental factors, such as viral infections, in immunogenetic predisposed individuals. Among these studies, the genetic predisposition underlying the mechanism of IA has been supported by some candidate gene associations in Western countries. Our study is the a sequence-based association study of IA, and we successfully identified common variants at the HLA region as well as rare variants in genes crucial for nerve and muscle that can contribute to the genetic basis of IA. The locus 6p21.32 was the only region with genome-wide

Table 2. The association results of 6 rare variants

Variant	Position	Gene	EA	Discovery (WES)		Validation (SANGER)		OR (combined)	p (combined)
				Variant carriers (subjects, n = 100)	Variant carriers (controls, n = 313)	Variant carriers (subjects, n = 230)	Variant carriers (controls, n = 1,760)		
rs142716067	chr7: 28,848,865	*CREB5*	T	3 (3.0%)	0 (0%)	4 (1.7%)	7 (0.40%)	6.38	1.26×10^{-3}
rs770625121	chr3: 138,183,253	*ESYT3*	T	2 (2.0%)	0 (0%)	3 (1.3%)	4 (0.23%)	7.95	3.72×10^{-3}
rs190743128	chr2: 11,925,128	*LPIN1*	G	3 (3.0%)	0 (0%)	2 (0.87%)	5 (0.28%)	6.37	6.57×10^{-3}
rs139010818	chr10: 102,056,033	*PKD2L1*	C	2 (2.0%)	0 (0%)	0 (0%)	0 (0%)	Inf	1.90×10^{-2}
rs762714174	chr2: 11,960,558	*LPIN1*	C	2 (2.0%)	0 (0%)	0 (0%)	1 (0.058%)	12.6	5.20×10^{-2}
rs140959397	chr3: 138,181,031	*ESYT3*	T	3 (3.0%)	0 (0%)	1 (0.43%)	19 (1.1%)	1.33	5.45×10^{-1}

A total of six recurrent IA rare variants in seven genes identified by the gene-based analysis were included in the validation. EA, effect allele; EAF, effect allele frequency; OR, odds ratio.

significance in the previous European GWASs.[9,36] Consistently, our results supported that 6p21.32 in the HLA region should be a locus enriched with many common variants associated with IA and emphasized the importance of immune-mediated processes in the etiology of IA, though no signals reached genome-wide significance ($p < 5 \times 10^{-8}$) in our array-based IA GWAS because of the relatively small number of affected individuals involved. However, population heterogeneity for the classical HLA genes was seen in previous studies[36–38] as well as in our study. The strongest signals identified by recent European studies were located at *HLA-DQ*; however, none of the *HLA-DQ* variants reached the significance criteria we defined. Instead, the strongest HLA-related SNP (rs1126511) in Chinese affected individuals was located at *HLA-DPB1* and changed the amino acid of *HLA-DPB1*. Those results indicate diverse genetic mechanisms underlying the population heterogeneity.

Another common variant, rs1705003, can affect the expression of *CUTA*, a copper-related protein, which can modulate b-amyloid and trigger a cascade of neurodegenerative steps and contribute to the development of Alzheimer disease (MIM: 104300).[39] Interestingly, the variant frequency is diverse across the population and is the highest in Ashkenazi Jews,[40] a group of individuals who have a disproportionately high prevalence of several autosomal-recessive genetic disorders, including multiple diseases affecting the nervous system (e.g., Canavan disease [MIM: 271900], familial dysautonomia [MIM: 223900], and Tay-Sachs disease [MIM: 272800]).[41] Thus, the variant may contribute to the genetic basis of the neurodegenerative process and results in deranged peristalsis and loss of the lower esophageal sphincter.[42] It is noteworthy that the function of the common variants in the HLA region could affect the target genes or transcripts in complicated ways. Since rs1705003 was associated with expression of the *CUTA* transcript (ENST00000488034) but it did not change the amino acid of the transcript, the amino acid-modified variants may also influence the expression of other transcripts by affecting the transcriptional process. Thus, the causal transcript and genetic mechanism will be of interest in further studies. In addition, we observed regulatory annotations on the linked variants of these amino acid-modified variants. Thus, a fine-mapping study with a large sample size followed by additional experiments may help illuminate the mechanism underlying the associations between variants and IA risk.

Beyond the common variants mentioned above, the gene-based analysis demonstrated the importance of rare variants in those genes that showed nervous system and muscle phenotypes in knock-out mice, further supporting the neurological and muscular basis of IA. Interestingly, most of the rare variants of these genes did not truncate the proteins and may not affect the function independently, but they were prone to co-inheritance in the IA-affected individuals and therefore may exert an effect on IA collaboratively. A similar combination model of missense variants was recently validated in congenital heart diseases (MIM: 600001) by CRISPR-Cas9 knock-in mice.[43] Thus, the rare missense variants should not be ignored in the study of IA and the genetic mechanisms underlying the variants are of interest for further investigation.

Among these variants, we successfully validated three variants in an independent dataset, which pinpointed three genes (*CREB5*, *ESYT3*, and *LPIN1*). *CREB5* belongs to the CRE (cAMP response element) binding protein family and is associated with anxiety (MIM: 607834) and airflow obstruction (MIM: 606963) by GWAS.[44,45] The *CREB* knock-out mice displayed an anxiety-like behavior,

suggesting that *CREB* may play a role in neurogenesis.[46–48] The rare variants of *ESYT3* (extended synaptotagmin 3) truncated its synaptotagmin-like domain (Figure S12). A previous study has reported that the loss[49] of *ESYT3* in mice was associated with an inhibition of behavioral or locomotor activity. *LPIN1* (lipin-1), expressed dominantly in nerve and muscle (Figure S13), plays a crucial role in lipid metabolism in multiple cell types, affecting processes ranging from myelin membrane biosynthesis to axo-glial interactions.[50,51] The deleterious variants in *LPIN1* in mice result in severe peripheral neuropathy (MIM: 300614) with transitory hindlimb paralysis.[51,52] The loss-of-function variants have also been identified in consanguineous families with recurrent acute myoglobinuria (MIM: 268200).[53] We noticed that rs762714174 (GenBank: NC_000002.11:g.11960558G>C) in *LPIN1* and rs140959397 (GenBank: NC_000003.11:g.138181031C>T) in *ESYT3* were not validated in the replication study. Although we did not observe mutated allele of rs762714174 in the affected individuals in the validation cohort, the frequency in the control subjects was extremely low (Table 2). Thus, the association between rs762714174 and IA still warranted investigation in a future study with expanded numbers of cases of IA. On the contrary, rs140959397 was not rare in the replication controls though the CADD score indicated its potential pathogenicity. Thus, large and ancestry-matched controls in the discovery cohort and an improved functional evaluation may help increase the validation rate in the study of rare variants.

In sum, we identified and successfully validated two common variants and three rare variants associated with IA in immunologic and neurological genes. Importantly, our results suggested that rare functional variants may contribute to IA with a combined effect, which provides new insight into the etiology of IA.

Data and code availability

The datasets of genotype information and the code supporting the current study are available from the corresponding author on request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.06.004.

Acknowledgments

This study was supported by grants from the National Natural Science Foundation of China (81873552, 81670483, and 81470811), Shanghai Rising-Star Program (19QA1401900), Major Project of Shanghai Municipal Science and Technology Committee (18ZR1406700), and the Outstanding Young Doctor Training Project of Shanghai Municipal Commission of Health and Family Planning (2017YQ026).

Declaration of Interests

The authors declare no competing interests.

Received: September 3, 2020
Accepted: June 2, 2021
Published: June 30, 2021

Web resources

Combined Annotation Dependent Depletion (CADD), https://cadd.gs.washington.edu/score

FastQC, https://www.bioinformatics.babraham.ac.uk/projects/fastqc

GenBank, https://www.ncbi.nlm.nih.gov/genbank/

Genotype-Tissue Expression (GTEx) data portal, https://www.gtexportal.org/home/datasets

Mouse Genome Informatics (MGI), http://www.informatics.jax.org/

OMIM, https://www.omim.org/

Picard, http://broadinstitute.github.io/picard

UCSC, http://genome.ucsc.edu/

VerifyBamID, https://github.com/statgen/verifyBamID

References

1. Harvey, P.R., Thomas, T., Chandan, J.S., Mytton, J., Coupland, B., Bhala, N., Evison, F., Patel, P., Nirantharakumar, K., and Trudgill, N.J. (2019). Incidence, morbidity and mortality of patients with achalasia in England: findings from a study of nationwide hospital and primary care data. Gut *68*, 790–795.
2. Boeckxstaens, G.E., Zaninotto, G., and Richter, J.E. (2014). Achalasia. Lancet *383*, 83–93.
3. Liu, Z.Q., Chen, W.F., Wang, Y., Xu, X.Y., Zeng, Y.G., Lee Dillon, D., Cheng, J., Xu, M.D., Zhong, Y.S., Zhang, Y.Q., et al. (2019). Mast cell infiltration associated with loss of interstitial cells of Cajal and neuronal degeneration in achalasia. Neurogastroenterol. Motil. *31*, e13565.
4. Gockel, H.R., Schumacher, J., Gockel, I., Lang, H., Haaf, T., and Nöthen, M.M. (2010). Achalasia: will genetic studies provide insights? Hum. Genet. *128*, 353–364.
5. Furuzawa-Carballeda, J., Torres-Landa, S., Valdovinos, M.A., Coss-Adame, E., Martín Del Campo, L.A., and Torres-Villalobos, G. (2016). New insights into the pathophysiology of achalasia and implications for future treatment. World J. Gastroenterol. *22*, 7892–7907.
6. Alahdab, Y.O., Eren, F., Giral, A., Gunduz, F., Kedrah, A.E., Atug, O., Yilmaz, Y., Kalayci, O., and Kalayci, C. (2012). Preliminary evidence of an association between the functional c-kit rs6554199 polymorphism and achalasia in a Turkish population. Neurogastroenterol. Motil. *24*, 27–30.
7. Nuñez, C., García-González, M.A., Santiago, J.L., Benito, M.S., Mearín, F., de la Concha, E.G., de la Serna, J.P., de León, A.R., Urcelay, E., and Vigo, A.G. (2011). Association of IL10 promoter polymorphisms with idiopathic achalasia. Hum. Immunol. *72*, 749–752.
8. de León, A.R., de la Serna, J.P., Santiago, J.L., Sevilla, C., Fernández-Arquero, M., de la Concha, E.G., Nuñez, C., Urcelay, E., and Vigo, A.G. (2010). Association between idiopathic achalasia and IL23R gene. Neurogastroenterol. Motil. *22*, 734–738, e218.
9. Gockel, I., Becker, J., Wouters, M.M., Niebisch, S., Gockel, H.R., Hess, T., Ramonet, D., Zimmermann, J., Vigo, A.G., Trynka, G., et al. (2014). Common variants in the HLA-DQ

region confer susceptibility to idiopathic achalasia. Nat. Genet. *46*, 901–904.
10. Santiago, J.L., de la Concha, E.G., de la Serna, J.P., Sevilla, C., Urcelay, E., and de León, A.R. (2012). Lack of association between the functional c-kit rs6554199 polymorphism and achalasia in a Spanish population. Hum. Immunol. *73*, 1207–1209.
11. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
12. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
13. Xie, C., Yeo, Z.X., Wong, M., Piper, J., Long, T., Kirkness, E.F., Biggs, W.H., Bloom, K., Spellman, S., Vierra-Green, C., et al. (2017). Fast and accurate HLA typing from short-read next-generation sequence data with xHLA. Proc. Natl. Acad. Sci. USA *114*, 8059–8064.
14. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
15. Pedersen, B.S., Layer, R.M., and Quinlan, A.R. (2016). Vcfanno: fast, flexible annotation of genetic variants. Genome Biol. *17*, 118.
16. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
17. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
18. Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., Richardson, J.E.; and Mouse Genome Database Group (2019). Mouse Genome Database (MGD) 2019. Nucleic Acids Res. *47* (D1), D801–D806.
19. Dai, J., Lv, J., Zhu, M., Wang, Y., Qin, N., Ma, H., He, Y.Q., Zhang, R., Tan, W., Fan, J., et al. (2019). Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. Lancet Respir. Med. *7*, 881–891.
20. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods *10*, 5–6.
21. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
22. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. *39*, 906–913.
23. Tullio-Pelet, A., Salomon, R., Hadj-Rabia, S., Mugnier, C., de Laet, M.H., Chaouachi, B., Bakiri, F., Brottier, P., Cattolico, L., Penet, C., et al. (2000). Mutant WD-repeat protein in triple-A syndrome. Nat. Genet. *26*, 332–335.
24. Handschug, K., Sperling, S., Yoon, S.J., Hennig, S., Clark, A.J., and Huebner, A. (2001). Triple A syndrome is caused by mutations in AAAS, a new WD-repeat protein gene. Hum. Mol. Genet. *10*, 283–290.
25. Roucher-Boulez, F., Brac de la Perriere, A., Jacquez, A., Chau, D., Guignat, L., Vial, C., Morel, Y., Nicolino, M., Raverot, G., and Pugeat, M. (2018). Triple-A syndrome: a wide spectrum of adrenal dysfunction. Eur. J. Endocrinol. *178*, 199–207.
26. Papageorgiou, L., Mimidis, K., Katsani, K.R., and Fakis, G. (2013). The genetic basis of triple A (Allgrove) syndrome in a Greek family. Gene *512*, 505–509.
27. Palka, C., Giuliani, R., Brancati, F., Mohn, A., Di Muzio, A., Calabrese, O., Huebner, A., De Grandis, D., Chiarelli, F., Ferlini, A., and Stuppia, L. (2010). Two Italian patients with novel AAAS gene mutation expand allelic and phenotypic spectrum of triple A (Allgrove) syndrome. Clin. Genet. *77*, 298–301.
28. Halpern, A.L., Torphy, R.J., McCarter, M.D., Sciotto, C.G., Glode, L.M., and Robinson, W.A. (2019). A familial germline mutation in KIT associated with achalasia, mastocytosis and gastrointestinal stromal tumors shows response to kinase inhibitors. Cancer Genet. *233-234*, 1–6.
29. Koehler, K., Milev, M.P., Prematilake, K., Reschke, F., Kutzner, S., Jühlen, R., Landgraf, D., Utine, E., Hazan, F., Diniz, G., et al. (2017). A novel *TRAPPC11* mutation in two Turkish families associated with cerebral atrophy, global retardation, scoliosis, achalasia and alacrima. J. Med. Genet. *54*, 176–185.
30. Busch, A., Žarković, M., Lowe, C., Jankofsky, M., Ganschow, R., Buers, I., Kurth, I., Reutter, H., Rutsch, F., and Hübner, C.A. (2017). Mutations in CRLF1 cause familial achalasia. Clin. Genet. *92*, 104–108.
31. van der Weyden, L., Happerfield, L., Arends, M.J., and Adams, D.J. (2009). Megaoesophagus in Rassf1a-null mice. Int. J. Exp. Pathol. *90*, 101–108.
32. Sivarao, D.V., Mashimo, H.L., Thatte, H.S., and Goyal, R.K. (2001). Lower esophageal sphincter is achalasic in nNOS(-/-) and hypotensive in W/W(v) mutant mice. Gastroenterology *121*, 34–42.
33. Taketomi, T., Yoshiga, D., Taniguchi, K., Kobayashi, T., Nonami, A., Kato, R., Sasaki, M., Sasaki, A., Ishibashi, H., Moriyama, M., et al. (2005). Loss of mammalian Sprouty2 leads to enteric neuronal hyperplasia and esophageal achalasia. Nat. Neurosci. *8*, 855–857.
34. Quang, D., Chen, Y., and Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics *31*, 761–763.
35. Shihab, H.A., Gough, J., Mort, M., Cooper, D.N., Day, I.N., and Gaunt, T.R. (2014). Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum. Genomics *8*, 11.
36. Becker, J., Haas, S.L., Mokrowiecka, A., Wasielica-Berger, J., Ateeb, Z., Bister, J., Elbe, P., Kowalski, M., Gawron-Kiszka, M., Majewski, M., et al. (2016). The HLA-DQbeta1 insertion is a strong achalasia risk factor and displays a geospatial north-south gradient among Europeans. Eur. J. Hum. Genet. *24*, 1228–1231.
37. Degenhardt, F., Wendorff, M., Wittig, M., Ellinghaus, E., Datta, L.W., Schembri, J., Ng, S.C., Rosati, E., Hübenthal, M., Ellinghaus, D., et al. (2019). Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. Hum. Mol. Genet. *28*, 2078–2092.
38. Furuzawa-Carballeda, J., Zuñiga, J., Hernández-Zaragoza, D.I., Barquera, R., Marques-García, E., Jiménez-Alvarez, L., Cruz-

Lagunas, A., Ramírez, G., Regino, N.E., Espinosa-Soto, R., et al. (2018). An original Eurasian haplotype, HLA-DRB1*14:54-DQB1*05:03, influences the susceptibility to idiopathic achalasia. PLoS ONE *13*, e0201676.

39. Hou, P., Liu, G., Zhao, Y., Shi, Z., Zheng, Q., Bu, G., Xu, H., and Zhang, Y.W. (2015). Role of copper and the copper-related protein CUTA in mediating APP processing and Aβ generation. Neurobiol. Aging *36*, 1310–1315.
40. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., and Alföldi, J. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-offunction intolerance across human protein-coding genes. bioRxiv. https://doi.org/10.1101/531210.
41. Weinstein, L.B. (2007). Selected genetic disorders affecting Ashkenazi Jewish families. Fam. Community Health *30*, 50–62.
42. Tuason, J., and Inoue, H. (2017). Current status of achalasia management: a review on diagnosis and treatment. J. Gastroenterol. *52*, 401–406.
43. Gifford, C.A., Ranade, S.S., Samarakoon, R., Salunga, H.T., de Soysa, T.Y., Huang, Y., Zhou, P., Elfenbein, A., Wyman, S.K., Bui, Y.K., et al. (2019). Oligogenic inheritance of a human heart disease involving a genetic modifier. Science *364*, 865–870.
44. Schosser, A., Butler, A.W., Uher, R., Ng, M.Y., Cohen-Woods, S., Craddock, N., Owen, M.J., Korszun, A., Gill, M., Rice, J., et al. (2013). Genome-wide association study of co-occurring anxiety in major depression. World J. Biol. Psychiatry *14*, 611–621.
45. Wilk, J.B., Shrine, N.R., Loehr, L.R., Zhao, J.H., Manichaikul, A., Lopez, L.M., Smith, A.V., Heckbert, S.R., Smolonska, J., Tang, W., et al. (2012). Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. Am. J. Respir. Crit. Care Med. *186*, 622–632.
46. Nakagawa, S., Kim, J.E., Lee, R., Chen, J., Fujioka, T., Malberg, J., Tsuji, S., and Duman, R.S. (2002). Localization of phosphorylated cAMP response element-binding protein in immature neurons of adult hippocampus. J. Neurosci. *22*, 9868–9876.
47. Gur, T.L., Conti, A.C., Holden, J., Bechtholt, A.J., Hill, T.E., Lucki, I., Malberg, J.E., and Blendy, J.A. (2007). cAMP response element-binding protein deficiency allows for increased neurogenesis and a rapid onset of antidepressant response. J. Neurosci. *27*, 7860–7868.
48. Gundersen, B.B., Briand, L.A., Onksen, J.L., Lelay, J., Kaestner, K.H., and Blendy, J.A. (2013). Increased hippocampal neurogenesis and accelerated response to antidepressants in mice with specific deletion of CREB in the hippocampus: role of cAMP response-element modulator τ. J. Neurosci. *33*, 13673–13685.
49. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al.; International Mouse Phenotyping Consortium; Jackson Laboratory; Infrastructure Nationale PHENOMIN, Institut Clinique de la Souris (ICS); Charles River Laboratories; MRC Harwell; Toronto Centre for Phenogenomics; Wellcome Trust Sanger Institute; and RIKEN BioResource Center (2016). High-throughput discovery of novel developmental phenotypes. Nature *537*, 508–514.
50. Finck, B.N., Gropler, M.C., Chen, Z., Leone, T.C., Croce, M.A., Harris, T.E., Lawrence, J.C., Jr., and Kelly, D.P. (2006). Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway. Cell Metab. *4*, 199–210.
51. Nadra, K., de Preux Charles, A.S., Médard, J.J., Hendriks, W.T., Han, G.S., Grès, S., Carman, G.M., Saulnier-Blache, J.S., Verheijen, M.H., and Chrast, R. (2008). Phosphatidic acid mediates demyelination in Lpin1 mutant mice. Genes Dev. *22*, 1647–1661.
52. Douglas, D.S., Moran, J.L., Bermingham, J.R., Jr., Chen, X.J., Brindley, D.N., Soliven, B., Beier, D.R., and Popko, B. (2009). Concurrent Lpin1 and Nrcam mouse mutations result in severe peripheral neuropathy with transitory hindlimb paralysis. J. Neurosci. *29*, 12089–12100.
53. Zeharia, A., Shaag, A., Houtkooper, R.H., Hindi, T., de Lonlay, P., Erez, G., Hubert, L., Saada, A., de Keyzer, Y., Eshel, G., et al. (2008). Mutations in LPIN1 cause recurrent acute myoglobinuria in childhood. Am. J. Hum. Genet. *83*, 489–494.