Transcriptome-wide association study identifies novel susceptibility genes contributing to hearing impairment

Chengyong Xie
Medical College of Guizhou University
Yuguang Niu
the First Medical Center of General Hospital of PLA
Jie Ping
Beijing Institute of Radiation Medicine
Qian Zhang
Medical College of Guizhou University
Zheng Zhang
Beijing Institute of Radiation Medicine
Yahui Wang
Beijing Institute of Radiation Medicine
Chenning Yang
Beijing Institute of Radiation Medicine
Gangqiao Zhou
Beijing Institute of Radiation Medicine
Yuanfeng Li (liyf_snp@163.com)
Beijing Institute of Radiation Medicine

Research Article

Keywords: Transcriptome-wide association study, Hearing impairment, ARL6IP6, TMEM127

Posted Date: April 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-465737/v1

Abstract

Hearing impairment (HI) is a growing public health issue of global concern, which may lead to decreased work efficiency, social withdrawal and depression. However, the knowledge of the genetic basis of HI is limited. We conducted a transcriptome-wide association study (TWAS) on risk of HI by combining an European cohort (578 cases and 583 controls) with a Chinese cohort (89 cases and 209 controls). We identified that 46 genes show transcriptome-wide significant associations with the HI risk in the meta-analyses of the European and the Chinese cohorts (all $P < 3.96 \times 10^{-6}$). Of these 46 candidate genes, four genes (*ARL6IP6*, *TMEM127*, *TOMM7* and *JAM3*) showed nominal evidence of association in both cohorts (all $P < 0.05$). Among these four genes, two ones (*ARL6IP6* and *TMEM127*) revealed strong, colocalized, and potentially causal associations with HI. Pathway enrichment analyses showed that the ciliary landscape ($P = 5.55 \times 10^{-4}$), cholesterol biosynthesis ($P = 7.09 \times 10^{-3}$), glycerophospholipid metabolism ($P = 9.65 \times 10^{-3}$) and IL-4 ($P = 1.11 \times 10^{-2}$) signaling pathways are significantly associated with HI. Our findings identified several novel susceptibility genes contributing to HI, and emphasized the power of TWAS in identifying putatively causal genes for common diseases.

Introduction

Hearing impairment (HI) is the most common sensorineural disorder around the world. It is estimated that HI is the fourth leading cause of disability, which affects approximately 1% of the population worldwide[1,2]. HI may lead to decreased work efficiency, social withdrawal and depression, therefore seriously affecting the life quality of humans, and causing huge economic losses to society.

It is commonly agreed that both genetic and environmental factors play important roles in the development of HI[3]. The known environmental risk factors of HI include noise, ototoxic drugs, ear infections and so on[2]. In addition to these environmental risk factors, genetic factors were also strongly suggested to be involved in HI by twin studies[4,5]. Recently, using genome-wide association studies (GWASs), we and others have identified a number of single nucleotide polymorphisms (SNPs) that were significantly associated with HI, such as rs7598759 (*NCL*), rs161927 (*GRM7*), rs2687481 (*GRM8*), rs681524 (*SIK3*), rs78043697 (*PCDH20*), rs7032430 (*SLC28A3*), rs9574464 (*DCLK1*), rs10815873 (*PTPRD*), rs898967 (*CMIP*), rs4932196 (*ISG20*) and rs5756795 (*TRIOBP*)[6-15].

Despite the significant success of GWASs in identifying loci that contribute to the genetic architecture of HI, a number of these identified loci reflect little direct biological relevance to this disease[16]. The transcriptome-wide association studies (TWASs), which integrated the GWAS data and the expression quantitative trait loci (eQTL) data, has emerged as a valuable supplyment to the GWAS. TWASs examine the associations between gene expressions and traits, rather than the associations between SNPs and traits. Thus, TWASs allowed for more interpretable biologically relevant results[17]. By use of TWAS, a number of genes have been identified to be significantly associated with several types of common diseases, including attention deficit/hyperactivity disorder (ADHD) and calcific aortic valve stenosis (CAVS)[18-20].

To identify novel genes conferring susceptibility to HI, we here performed a TWAS by combining the individual-level genotype data of two HI cohorts from two previously published HI GWASs[13,21] with the whole blood eQTL data from Genotype-Tissue Expression (GTEx) (v8). We identified *ARL6IP6* on chromosome 2q23.3 and *TMEM127* on chromosome 2q11.2 as the candidate genes of HI. Our findings expanded our understanding of the genetic susceptibility to HI.

Results

Transcriptome-wide association analyses

To detect novel genes conferring susceptibility to HI, we conducted a TWAS using PrediXcan (**Fig. 1**). In this TWAS, we collected two publicly available GWAS datasets from a Chinese case-control population and an European case-control population, respectively (**Table 1**). The Chinese cohort included 298 individuals, consisting of 89 cases and 209 controls. The European cohort included 1,161 individuals, consisting of 578 cases and 583 controls. No outlier was present using the PCA (**Supplementary Fig. 1**). To ensure the robustness of follow-up analyses, we conducted a unified quality controls (Methods), and finally retained 3,830,463 SNPs in the Chinese cohort and 5,767,726 in the European cohort, respectively. We then imputed the gene expressions based on the genotypes of these SNPs using the eQTL data of 670 blood tissues from GTEx (v8) as reference. Thus, we generated the expression levels for a total of 12,624 genes.

Next, we carried out gene-phenotype association analyses for the two cohorts in a linear regression model, with adjustment for the sex, principal component (PC) 1 and PC2 from PCA. In the European cohort, a total of 171 genes reached the transcriptome-wide significance threshold of $P < 3.96 \times 10^{-6}$ (0.05/12,624) (**Supplementary Table 1**). Among these 171 genes, 90 ones were upregulated, and 81 ones were downregulated in HI cases. In the Chinese cohort, no gene reached the transcriptome-wide significance threshold, however, 231 genes showed nominal evidence of association ($P < 0.05$; **Supplementary Table 2**). Among these 231 genes, 121 ones were upregulated, and 110 ones were downregulated in HI cases.

We then conducted meta-analyses to combine the association results of the European cohort and the Chinese cohort. Manhattan plots were created to show the association and direction between genes and HI risk in the meta-analyses (**Fig. 2**). In the meta-analyses, we identified a total of 46 genes significantly associated with HI ($P < 3.96 \times 10^{-6}$, **Supplementary Table 3**). Among these 46 genes, four ones (*ARL6IP6*, *TMEM127*, *TOMM7* and *JAM3*) showed nominal evidence of association in both cohorts ($P < 0.05$), and the association directions for these 4 genes were consistent across the two cohorts (**Table 2**). Subsequent conditional analyses showed that the associations of these four genes were independent each other. Further, these four genes were outside of the loci that were reported to be significantly associated with HI (± 500 kb). These results suggested that these four genes may be the novel candidate susceptibility genes contributing to HI.

Four candidate susceptibility genes

ARL6IP6 (ADP Ribosylation Factor Like GTPase 6 Interacting Protein 6) gene, which is located on chromosome 2q23.3, reached the transcriptome-wide significance ($P_{TWAS} = 5.85 \times 10^{-8}$). The rs7589899 was the lead SNP at the *ARL6IP6* locus that was significantly associated with HI (odds ratio [OR] for T allele = 0.90,

95% confidence interval [CI] = 0.76-1.07, $P_{GWAS} = 7.70 \times 10^{-14}$). This SNP was located in the third intron of *ARL6IP6*. The risk allele [C] of this SNP was significantly associated with the high expression of *ARL6IP6* in whole blood ($P_{eQTL} = 0.011$; **Supplementary Figs. 2a** and **2b**). Formal Bayesian colocalization revealed a posterior probability (PP) of shared signals of 57.5% (**Fig. 3a**), providing evidence that the GWAS and eQTL signals may share the same variants at the *ARL6IP6* locus. Conditional analyses showed that the association between rs7589899 and HI was not significant after adjusting for the expression of *ARL6IP6* ($P_{GWAS} = 0.88$), indicating that *ARL6IP6* explains all of the signal at its locus. These results suggested that high expressions of *ARL6IP6* may increase the risk of HI.

TMEM127 (Transmembrane Protein 127) gene, which is located on chromosome 2q11.2, reached the transcriptome-wide significance ($P_{TWAS} = 1.78 \times 10^{-6}$). The rs607302 was the lead SNP at *TMEM127* locus that was most significantly associated with HI (OR = 1.12, 95% CI = 0.95-1.32, $P_{GWAS} = 1.01 \times 10^{-4}$). rs607302 is an intergenic variant. The blood eQTL rs607302-*TMEM127* indicated that the risk allele [G] for HI is associated with lower mRNA expression levels of *TMEM127* in whole blood (**Supplementary Figs. 2c** and **2d**). The formal Bayesian colocalization revealed a PP of shared signals of 75.5% (**Fig. 3b**), providing evidence that the GWAS and eQTL signals may share the same variants at the *TMEM127* locus. Conditional analyses showed that the association between rs607302 and HI was not significant after adjusting for the expression of *TMEM127* ($P_{GWAS} = 1$), indicating that *TMEM127* explains all of the signal at its locus. These results suggested that higher expressions of *TMEM127* decrease the risk of HI.

TOMM7 (Translocase Of Outer Mitochondrial Membrane 7) gene, which is located on chromosome 7p15.3, reached the transcriptome-wide significance ($P_{TWAS} = 2.87 \times 10^{-6}$). The rs12534523 was the lead GWAS SNP at *TOMM7* locus. It is normally associated with HI (OR = 0.79, 95% CI = 0.66-0.95, $P_{GWAS} = 2.02 \times 10^{-3}$), but is not associated with the expression of *TOMM7* ($P_{eQTL} = 0.11$; **Fig. 3c**, **Supplementary Figs. 2e** and **2f**). Formal Bayesian colocalization also revealed a PP of shared signals (PP4) of 1.29%, suggesting that the GWAS and eQTL signals may not share the same variants at the *TOMM7* locus. These results suggested that the association between *TOMM7* expression and HI may be a joint effect of multiple SNPs.

JAM3 (Junctional Adhesion Molecule 3) gene, which is located on chromosome 11q25, reached transcriptome-wide significance ($P_{TWAS} = 3.33 \times 10^{-6}$). rs10750560 was the lead GWAS SNP at *JAM3* locus that was normally associated with both HI (OR = 0.79, 95% CI = 0.66-0.95, $P_{GWAS} = 2.42 \times 10^{-3}$) and the expression of *JAM3* ($P_{eQTL} = 1.69 \times 10^{-2}$; **Fig. 3d**, **Supplementary Figs. 2g** and **2h**). Formal Bayesian colocalization revealed a PP of shared signals (PP4) of 0.76%, providing evidence that the GWAS and eQTL signals may not share the same variants at the *JAM3* locus. These results suggested that the association between *JAM3* expression and HI may be a joint effect of multiple SNPs.

Pathway enrichment

To understand the biologically relevant pathways involved in HI, pathway analyses were conducted using KEGG, Wiki and Reactome Pathways in the EnrichR web server[22,23]. We used the 46 significantly associated genes in the meta-analyses of the two cohorts ($P_{TWAS} < 3.96 \times 10^{-6}$). In this analysis, no gene sets reached the

significant threshold of $P < 2.59 \times 10^{-5}$ (0.05/1,927) accounting for multiple testing corrections. However, 12 candidate gene sets were identified to be nominally significantly associated with HI ($P < 0.05$; **Supplementary Table 4**). Intriguingly, four of these 12 candidate gene sets were biological plausibility in the development of HI[24-27], including ciliary landscape ($P = 5.55 \times 10^{-4}$, ranks the first), cholesterol biosynthesis ($P = 7.09 \times 10^{-3}$, ranks the third), glycerophospholipid metabolism ($P = 9.65 \times 10^{-3}$, ranks the fifth) and IL-4 signaling pathways ($P = 1.11 \times 10^{-2}$, ranks the sixth) (**Supplementary Table 4**). For example, the cilia on the inner and outer hair cells can convert mechanical deflection signals into electrochemical signals, and participate in the transmission of sound[28]. Cholesterol metabolic pathways were demonstrated to play important roles in recovery of hearing ability following noise-induced hearing loss[25]. IL-4 plays a key role in ototoxicity by activating the phosphorylation of STAT6 and promoting the production of inflammatory cytokines[27], which were involved in mediating cochlear damage during the acute stage of HI[27]. Glycerophospholipid metabolism was reported to be involved in the the regulation of voltage-gated K+ channel[26], which plays important roles in the inner ear and hearing loss. Taken together, using TWAS, we identified several pathways significantly associated with HI. However, the underlying mechanisms of these pathways in the development of HI require further investigation.

Phenome-wide association study

To identify the traits that were significantly associated or co-morbid with HI, and to provide new insights into HI etiology, a phenome-wide association study (pheWAS) was performed for the four significant genes (*ARL6IP6*, *TMEM127*, *TOMM7* and *JAM3*). We emphasized the top three phenotypes for each gene (**Supplementary Table 5**). *ARL6IP6* was strongly associated with the age at menarche ($P = 1.01 \times 10^{-11}$), estimated bone mineral density from heel ultrasounds ($P = 7.89 \times 10^{-11}$) and heel bone mineral density ($P = 8.87 \times 10^{-11}$). *TMEM127* was strongly associated with the monocyte percentage ($P = 1.95 \times 10^{-9}$), eosinophil percentage ($P = 2.57 \times 10^{-9}$) and monocyte percentage of white cells ($P = 5.89 \times 10^{-9}$). *TOMM7* was associated with the hemoglobin concentration ($P = 2.20 \times 10^{-6}$), type 2 diabetes ($P = 3.39 \times 10^{-6}$) and “Never eat eggs, dairy, wheat, sugar” ($P = 1.87 \times 10^{-5}$). *JAM3* was associated with the narcolepsy ($P = 2.03 \times 10^{-9}$), schizophrenia ($P = 9.23 \times 10^{-8}$) and waist-hip ratio ($P = 5.60 \times 10^{-5}$).

To further determine whether the pheWAS traits were genetically correlated with HI, and in which direction, we performed the genetic correlation analyses between HI and the 12 phenotypes from pheWAS[29] (**Supplementary Fig. 3**). Interestingly, we detected significant positive genetic associations between HI and two other diseases: the narcolepsy ($r = 0.164$, $P = 9.38 \times 10^{-7}$) and waist-to-hip ratio ($r = 0.114$, $P = 1.28 \times 10^{-5}$). In addition, we also found normally significant positive genetic associations between HI and type 2 diabetes ($r = 0.087$, $P = 7.31 \times 10^{-4}$), as well as “Never eat eggs, dairy, wheat, sugar” ($r = 0.290$, $P = 8.18 \times 10^{-3}$) and schizophrenia ($r = 0.070$, $P = 4.60 \times 10^{-3}$). On the contrary, we detected a normally significant negative genetic association between HI and the age at menarche ($r = -0.061$, $P = 1.70 \times 10^{-2}$). Several of these phenotypes were previously thought to be closely related to HI, thus reaffirming the genetic relevance[30-33]. For examples, higher waist-hip ratio and type 2 diabetes have been previously shown to be risk factors for HI[34-36]. In addition, there is a lot of overlapping symptoms in HI and neurological diseases, including schizophrenia[37]. Taken together, these results suggested that there are relationships between HI and several other traits.

Discussion

In the present study, we conducted a TWAS on the risk of HI. We identified four novel candidate genes *ARL6IP6* (2q23.3), *TMEM127* (2q11.2), *TOMM7* (7p15.3), *JAM3* (11q25) contributing to the genetic susceptibility to HI. We also showed that the ciliary landscape, cholesterol biosynthesis, glycerophospholipid metabolism and IL-4 signaling pathways are associated with HI. To our best knowledge, this is the first TWAS for the genetic susceptibility to risk of HI.

Recent GWASs have successfully identified several loci significantly associated with HI. However, the functional significance of these associations remain elusive due to the inability to fine-map to tissue-specific and tissue-relevant genes. In fact, GWAS loci for other complex traits are also difficult to interpret. Thus, the development of methods to prioritize causal genes of GWAS attracts the attentions of geneticists. TWAS is one of the methods to solve this problem, which integrates genotype, gene expression and phenotype to gain insights into the genetic basis of complex traits. In this study, using TWAS, we identified four genes associated with HI risk in the whole blood. Thus, our results demonstrated the power of TWAS.

ARL6IP6 is involved in DNA repair and apoptosis[38], which is mainly expressed in nucleus and plasma membrane. Previous GWAS identified that *ARL6IP6* is significantly associated with ischemic attacks[39]. Moreover, a homozygous truncating mutation in *ARL6IP6* was suggested to be the likely cause of cutis marmorata telangiectatica congenita (CMTC), which is associated with ischemic attacks[40]. Interestingly, evidence has shown that ischemic stroke is associated with HI. Previous studies have shown that the blood supply of the auditory system comes from the vertebral basal system, so a ischemic stroke in the vertebral basal circulation can present with acute HI[41]. In fact, the incidence of HI following ischemic stroke is 8.0%[42]. Thus, the finding that *ARL6IP6* is a potential HI susceptibility gene provided genetic evidence for this phenomenon.

TMEM127 encodes a transmembrane protein with 3 predicted transmembrane domains. Previous studies suggested that *TMEM127* is mainly expressed in the plasma membrane and endosome[43], and it dynamically associates with the endomembrane system[44]. *In vitro* assays indicate that *TMEM127* is a negative regulator of mammalian target of rapamycin (mTOR) signaling[44]. Mouse models showed that sustained mTOR activation disrupted the redox balance in neurosensory epithelium, thus causing early-onset death of cochlear hair cells and progressive HI[45]. Treatment with rapamycin reduced the activity of mTOR in cochlear hair cells, and reduced hair cells from injury *in vivo*[45]. Collectively, these observations suggested that *TMEM127* is a candidate gene for the genetic susceptibility to HI.

TOMM7 encodes a component of the protein translocase of the outer mitochondrial membrane (TOM) complex[46]. Previous studies have revealed that *TOMM7* was involved in the sorting and accumulation of the preproteins at the outer membrane, which played an important role in protein import[47]. *JAM3* encodes a member of the junctional adhesion molecule family, and is reported to be involved in the regulation of cell migration or polarization[48]. However, no evidence has ever shown that these two genes was involved in HI. Further studies are needed to reveal the underlying mechanisms of these two genes contributing to HI.

Several limitations of this study are worth mentioning. First, although this study included two independent cohorts, the sample size remains limited. Replications in additionally independent cohorts are needed. Secondly, the pleiotropy is statistically indistinguishable from truly causal genes. Thus, the susceptibility genes identified by TWAS may be not the truly causal genes. Experimental validations are needed.

In summary, our TWAS reveals novel genes contributing to the susceptibility to HI across the European cohort and the Chinese cohort. These findings expand our understanding of the genetic susceptibility to HI and might shed light on the prevention and clinical implications of the candidate genes in the treatment of HI.

Methods

Study participants

In the present study, we collected two independent datasets for GWAS of HI (**Table 1**). The first GWAS dataset was derived from a Chinese HI cohort, consisting of 89 cases and 209 controls[21]. The second GWAS dataset derived from an European cohort, consisting of 578 cases and 583 controls[13]. The genetic origins of the subjects in these two cohorts were confirmed by principal component analysis (PCA).

According to the World Health Organization, if the minimum threshold of hearing on either side of the ears of a subject is greater than 25 decibel (dB), this subject can be regarded as a case of HI[49]. According to this criteria, the Chinese cohort contains 89 cases and 209 controls. The mean age of the cases (23.8) is similar to that of controls (23.3) ($P > 0.05$), and all the cases and controls are males (**Table 1**). The European cohort contains 578 cases and 583 controls. The mean age of the cases (61.2) is similar to that of controls (61.4) ($P > 0.05$). The male/female ratios of the cases and controls are 0.98 (286/292) and 1.06 (300/283), respectively ($P > 0.05$; **Table 1**).

GWAS data of the Chinese cohort

This GWAS dataset was derived from our previous GWAS for HI in Chinese population, which was publicly available through the following link http://cbportal.org/pubfiles/JCMM-07-2020-200.zip. Detailed information on these subjects was described previously[21]. Briefly, this GWAS contained a total of 89 cases and 209 controls. All the subjects were recruited from a single factory at Bengbu city, Anhui province, China. The SNPs were genotyped using the Illumina Infinium Asian Screening Array-24 (v1.0), and imputed using IMPUTE2 software (v2.3.1). We performed SNP quality controls for further analyses. SNPs were excluded if they: (i) had a call rate < 90%; (ii) had a minor allele frequency (MAF) < 0.05; (iii) were not autosomal SNPs; or (iv) a P value < 1.0×10^{-4} in Hardy-Weinberg equilibrium (HWE) test. After filtering, a total of 3,830,463 SNPs were retained.

GWAS data of the European cohort

This GWAS dataset was derived from a GWAS for HI in European population, which was publicly available through the following link https://tgen.org/research/research

-divisions/neurogenomics/supplementary-data/gwas_polygenic_arhi_fransen_et_al.aspx. Detailed information on these subjects was described previously[13]. Briefly, this GWAS contained a total of 578 cases

and 583 controls. All the subjects were recruited from the Antwerp University Hospital at Antwerp city, Belgium[13]. The SNPs were genotyped using the Illumina CNV370 Quad Chip or Illumina HumanOmniExpress BeadChip (Illumina, Inc., San Diego, CA, USA), and imputed using IMPUTE2 software (V2.1.2). We performed the same quality controls as that used in the GWAS dataset of the Chinese cohort. After filtering, a total of 5,767,726 SNPs were included in this analysis.

Principal component analysis (PCA)

To quantify the population structure, principal component analysis (PCA) was implemented in EIGENSTRAT (v3.0). PCA was conducted on the subjects in these two cohorts of this study and additional 2,504 subjects from 1,000 Genomes Project, including 107 TSIs (Toscani in Italia), 107 IBSs (Iberian Population in Spain), 91 GBRs (British in England and Scotland), 99 FINs (Finnish in Finland), 99 CEUs (Utah Residents (CEPH) with Northern and Western European Ancestry), 93 CDXs (Chinese Dai in Xishuangbanna, China), 105 CHSs (Southern Han Chinese), 103 CHBs (Han Chinese in Beijing, China), 99 KHVs (Kinh in Ho Chi Minh City, Vietnam), and 104 JPTs (Japanese in Tokyo, Japan).

Transcriptome-wide association study

The TWAS was performed using the PrediXcan software[50]. PrediXcan is a gene-based association test software that prioritizes genes that are likely to be causal for the phenotype. PrediXcan firstly builds transcriptome expression prediction models based on expression quantitative trait loci (eQTL) data. Then, using the prediction models, PrediXcan imputes gene expressions at the individual level based on the genome-wide SNPs data. Finally, PrediXcan performs the association tests based on the imputed transcriptome levels[51]. In the present study, we used the prediction model from the blood tissues (n = 670) from GTEx (v8) database in the PredictDB data repository. We performed TWAS on the Chinese cohort and the European cohort separately. Meta-analyses were then performed to combine the association results of these two cohorts. In this study, a total of 12,624 genes were analyzed, leading to the transcriptome-wide significance threshold of $P = 3.96 \times 10^{-6}$ (0.05/12,624).

Bayesian colocalization

To test whether the leading variant of the GWAS and the eQTL signal was the same one, we used the COLOC (version 3.2-1) pack in R to perform colocalization analyses[52]. As the majority of gene expression datasets were generated on the basis of European-ancestry samples, colocalization analyses were based on the GWAS summary statistics from the European cohort and the whole blood eQTL data from European population from the GTEx (v8). COLOC tested for five hypotheses: H0, no eQTL and no GWAS association; H1, association with eQTL, but no GWAS; H2, association with GWAS, but no eQTL; H3, eQTL and GWAS association, but independent signals; and H4, shared eQTL and GWAS associations. The main interest is to assess whether the GWAS and eQTL signals are consistent with shared causal variants (*i.e.*, H4). The result of this procedure is five posterior probabilities (PP0, PP1, PP2, PP3 and PP4). In practice, a high posterior probability (PP4 > 50%) indicates that the GWAS and eQTL signals colocalize[53]. We used an R package, LocusCompareR, for visualization of colocalization events in local environments, which generates a combined plot with two locus-zoom plots (eQTL and GWAS in the same gene region) and a locus-compare scatter plot (eQTL -$\log_{10}$ (P) to

GWAS -$\log_{10}$ (*P*)). The figure could indicate whether the GWAS top locus is also the leading SNP in the eQTL result[54].

Pathway enrichment analyses

To explore the potential biologically relevant pathways involved in HI, we performed pathway enrichment analyses using the EnrichR web server (https://maayanlab.cloud/

Enrichr/)[22,23]. In this analyses, we used 46 significantly associated genes in the meta-analyses of the two cohorts ($P_{TWAS} < 3.96 \times 10^{-6}$). A total of 186 gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/), 242 gene sets from the WikiPathways (https://www.wikipathways.org/index.php/WikiPathways) and 1,499 gene sets from the Reactome database (http://www.reactome.org) were included in our analyses. We performed multiple testing corrections using the Benjamini-Hochberg (BH) method, and the false discovery rate (FDR) values of less than 0.05 were considered to be statistically significant. However, no gene set reached a significant threshold accounting for multiple testing correction (all FDR > 0.05).

Phenome-wide association studies

To check the associations between the four significantly associated genes in this study (*ARL6IP6*, *TMEM127*, *TOMM7* and *JAM3*) and a wide range of phenotypes in the UK Biobank, we performed a phenome-wide association study (pheWAS) using the publicly available data provided by GWAS Atlas (https://atlas.ctglab.nl)[29]. The database contains 4,756 GWAS from 473 unique studies across 3,302 unique traits. We used the colocalization procedure described above to test for shared causal associations between HI and other phenotypes. The top three phenotypes of each gene were considered to share causal associations with HI[55]. phenotypes were identified by pheWAS.

Genetic correlation of pheWAS traits

To determine the genetic relationship between HI and the phenotypes identified from pheWAS (n = 12), we used linkage disequilibrium score regression (LDSC) software (v1.0.1)[56]. The genetic correlations were calculated based on the GWAS summary statistics from GWASAtlas (https://atlas.ctglab.nl). Genetic correlation tests were performed 78 times (12 × 13 / 2). We set the significance threshold as $P = 6.41 \times 10^{-4}$ (0.05/78), based on the Bonferroni correction using the number of tests.

Statistical analyses

The gene-phenotype association analyses were performed in a linear regression model with adjustment for the sex, principal component (PC) 1 and PC2 from PCA. Meta-analyses of the association results generated from the Chinese cohort and European cohort were conducted to assess the pooled genetic effects using the R statistical package (version 4.0.3). Cochran's Q statistic was calculated to test the heterogeneity between groups. The *P*-values from the meta-analyses were corrected for multiple testing based on FDR correction[57].

Declarations

Ethics approval and consent to participate

This study was performed with the approval of the Medical Ethical Committee of Beijing Institute of Radiation Medicine (Beijing, China) and General Hospital of PLA (Beijing, China), and the research carried out in accordance with the guidelines of both committees. Written informed consent was obtained from each participant. The investigators were blind to the case/control status of subjects during all genotyping experiments.

Data availability

The datasets generated during and/or analysed during the current study are available for download on the journal website.

Acknowledgments

The authors thank all the patients participating in this study. This work was supported by grants from National Natural Science Foundation of China (No. 81730055).

Authors' contributions

G.Z. and Y.L. were the principal investigators who conceived the study and obtained financial supports. G.Z., Y.L., Y.N., and C.X. designed the study. C.X. analyzed the data. Z.Z., Q.Z., P.J. and C.Y. conducted sample selection and data management, Y.W. performed the statistical analyses, G.Z., Y.L. and C.X. interpreted the results, G.Z, Y.L. and C.X. drafted the manuscript. G.Z. approved the final version of the manuscript.

Competing interests

The authors declared no competing interests.

References

1 Olusanya, B. O., Davis, A. C. & Hoffman, H. J. Hearing loss: rising prevalence and impact. *Bull World Health Organ* **97**, 646-646A, doi:10.2471/BLT.19.224683 (2019).

2 Cunningham, L. L. & Tucci, D. L. Hearing Loss in Adults. *N Engl J Med* **377**, 2465-2473, doi:10.1056/NEJMra1616601 (2017).

3 Korver, A. M. *et al.* Congenital hearing loss. *Nat Rev Dis Primers* **3**, 16094, doi:10.1038/nrdp.2016.94 (2017).

4 Bogo, R. *et al.* Prevalence, Incidence Proportion, and Heritability for Tinnitus: A Longitudinal Twin Study. *Ear Hear* **38**, 292-300, doi:10.1097/AUD.0000000000000397 (2017).

5 Duan, H. *et al.* Heritability of Age-Related Hearing Loss in Middle-Aged and Elderly Chinese: A Population-Based Twin Study. *Ear Hear* **40**, 253-259, doi:10.1097/AUD.0000000000000610 (2019).

6 Wells, H. R. R. *et al.* GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *Am J Hum Genet* **105**, 788-802, doi:10.1016/j.ajhg.2019.09.008 (2019).

7 Vuckovic, D. *et al.* Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. *Hum Mol Genet* **24**, 5655-5664, doi:10.1093/hmg/ddv279 (2015).

8 Niu, Y. *et al.* Genome-wide association study identifies 7q11.22 and 7q36.3 associated with noise-induced hearing loss among Chinese population. *J Cell Mol Med* **25**, 411-420, doi:10.1111/jcmm.16094 (2021).

9 Hoffmann, T. J. *et al.* A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. *PLoS Genet* **12**, e1006371, doi:10.1371/journal.pgen.1006371 (2016).

10 Girotto, G. *et al.* Hearing function and thresholds: a genome-wide association study in European isolated populations identifies new loci and pathways. *J Med Genet* **48**, 369-374, doi:10.1136/jmg.2010.088310 (2011).

11 Nagtegaal, A. P. *et al.* Genome-wide association meta-analysis identifies five novel loci for age-related hearing impairment. *Sci Rep* **9**, 15192, doi:10.1038/s41598-019-51630-x (2019).

12 Van Laer, L. *et al.* A genome-wide association study for age-related hearing impairment in the Saami. *Eur J Hum Genet* **18**, 685-693, doi:10.1038/ejhg.2009.234 (2010).

13 Fransen, E. *et al.* Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *Eur J Hum Genet* **23**, 110-115, doi:10.1038/ejhg.2014.56 (2015).

14 Wheeler, H. E. *et al.* Variants in WFS1 and Other Mendelian Deafness Genes Are Associated with Cisplatin-Associated Ototoxicity. *Clin Cancer Res* **23**, 3325-3333, doi:10.1158/1078-0432.CCR-16-2809 (2017).

15 Ali Mosrati, M. *et al.* Genome-wide analysis reveals a novel autosomal-recessive hearing loss locus DFNB80 on chromosome 2p16.1-p21. *J Hum Genet* **58**, 98-101, doi:10.1038/jhg.2012.141 (2013).

16 Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467-484, doi:10.1038/s41576-019-0127-1 (2019).

17 Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-252, doi:10.1038/ng.3506 (2016).

18 Hammerschlag, A. R. *et al.* Refining Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder Genetic Loci by Integrating Summary Data From Genome-wide Association, Gene Expression, and DNA Methylation Studies. *Biol Psychiatry* **88**, 470-479, doi:10.1016/j.biopsych.2020.05.002 (2020).

19 Theriault, S. *et al.* A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. *Nat Commun* **9**, 988, doi:10.1038/s41467-018-03260-6 (2018).

20 Liao, C. *et al.* Transcriptome-wide association study of attention deficit hyperactivity disorder identifies associated genes and phenotypes. *Nat Commun* **10**, 4450, doi:10.1038/s41467-019-12450-9 (2019).

21 Niu, Y. *et al.* Genome-wide association study identifies 7q11.22 and 7q36.3 associated with noise-induced hearing loss among Chinese population. *J Cell Mol Med*, doi:10.1111/jcmm.16094 (2020).

22 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).

23 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).

24 Baker, K. & Beales, P. L. Making sense of cilia in disease: the human ciliopathies. *Am J Med Genet C Semin Med Genet* **151C**, 281-295, doi:10.1002/ajmg.c.30231 (2009).

25 Sai, N. *et al.* Involvement of Cholesterol Metabolic Pathways in Recovery from Noise-Induced Hearing Loss. *Neural Plast* **2020**, 6235948, doi:10.1155/2020/6235948 (2020).

26 Taylor, K. C. & Sanders, C. R. Regulation of KCNQ/Kv7 family voltage-gated K(+) channels by lipids. *Biochim Biophys Acta Biomembr* **1859**, 586-597, doi:10.1016/j.bbamem.2016.10.023 (2017).

27 Kim, H. J. *et al.* Cisplatin ototoxicity involves cytokines and STAT6 signaling network. *Cell Res* **21**, 944-956, doi:10.1038/cr.2011.27 (2011).

28 Fettiplace, R. Hair Cell Transduction, Tuning, and Synaptic Transmission in the Mammalian Cochlea. *Compr Physiol* **7**, 1197-1227, doi:10.1002/cphy.c160049 (2017).

29 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).

30 Sindura, K. P. & Banerjee, M. An Immunological Perspective to Non-syndromic Sensorineural Hearing Loss. *Front Immunol* **10**, 2848, doi:10.3389/fimmu.2019.02848 (2019).

31 Baradaranfar, M. *et al.* The Role of Immunological Factors on Sudden Sensoryneural Hearing Loss. *Iran J Otorhinolaryngol* **30**, 219-223 (2018).

32 Blazer, D. G. & Tucci, D. L. Hearing loss and psychiatric disorders: a review. *Psychol Med* **49**, 891-897, doi:10.1017/S0033291718003409 (2019).

33 Blazer, D. G. Hearing Loss: The Silent Risk for Psychiatric Disorders in Late Life. *Psychiatr Clin North Am* **41**, 19-27, doi:10.1016/j.psc.2017.10.002 (2018).

34 Aladag, I., Eyibilen, A., Guven, M., Atis, O. & Erkorkmaz, U. Role of oxidative stress in hearing impairment in patients with type two diabetes mellitus. *J Laryngol Otol* **123**, 957-963, doi:10.1017/S0022215109004502 (2009).

35 Curhan, S. G., Eavey, R., Wang, M., Stampfer, M. J. & Curhan, G. C. Body mass index, waist circumference, physical activity, and risk of hearing loss in women. *Am J Med* **126**, 1142 e1141-1148, doi:10.1016/j.amjmed.2013.04.026 (2013).

36 Cruickshanks, K. J. *et al.* Smoking, central adiposity, and poor glycemic control increase risk of hearing impairment. *J Am Geriatr Soc* **63**, 918-924, doi:10.1111/jgs.13401 (2015).

37 Kancherla, V., Van Naarden Braun, K. & Yeargin-Allsopp, M. Childhood vision impairment, hearing loss and co-occurring autism spectrum disorder. *Disabil Health J* **6**, 333-342, doi:10.1016/j.dhjo.2013.05.003 (2013).

38 Cox, L., Vanderwall, D. K., Parkinson, K. C., Sweat, A. & Isom, S. C. Expression profiles of select genes in cumulus-oocyte complexes from young and aged mares. *Reprod Fertil Dev* **27**, 914-924, doi:10.1071/RD14446 (2015).

39 Cheng, Y. C. *et al.* Genome-wide association analysis of ischemic stroke in young adults. *G3 (Bethesda)* **1**, 505-514, doi:10.1534/g3.111.001164 (2011).

40 Abumansour, I. S. *et al.* ARL6IP6, a susceptibility locus for ischemic stroke, is mutated in a patient with syndromic Cutis Marmorata Telangiectatica Congenita. *Hum Genet* **134**, 815-822, doi:10.1007/s00439-015-1561-6 (2015).

41 Kim, H. A., Yi, H. A. & Lee, H. Recent Advances in Cerebellar Ischemic Stroke Syndromes Causing Vertigo and Hearing Loss. *Cerebellum* **15**, 781-788, doi:10.1007/s12311-015-0745-x (2016).

42 Lee, H. & Baloh, R. W. Sudden deafness in vertebrobasilar ischemia: clinical features, vascular topographical patterns and long-term outcome. *J Neurol Sci* **228**, 99-104, doi:10.1016/j.jns.2004.10.016 (2005).

43 Yao, L. *et al.* Spectrum and prevalence of FP/TMEM127 gene mutations in pheochromocytomas and paragangliomas. *JAMA* **304**, 2611-2619, doi:10.1001/jama.2010.1830 (2010).

44 Qin, Y. *et al.* Germline mutations in TMEM127 confer susceptibility to pheochromocytoma. *Nat Genet* **42**, 229-233, doi:10.1038/ng.533 (2010).

45 Fu, X. *et al.* Tuberous sclerosis complex-mediated mTORC1 overactivation promotes age-related hearing loss. *J Clin Invest* **128**, 4938-4955, doi:10.1172/JCI98058 (2018).

46 Hasson, S. A. *et al.* High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291-295, doi:10.1038/nature12748 (2013).

47 Honlinger, A. *et al.* Tom7 modulates the dynamics of the mitochondrial outer membrane translocase and plays a pathway-related role in protein import. *EMBO J* **15**, 2125-2137 (1996).

48 Aramsangtienchai, P., Spiegelman, N. A., Cao, J. & Lin, H. S-Palmitoylation of Junctional Adhesion Molecule C Regulates Its Tight Junction Localization and Cell Migration. *J Biol Chem* **292**, 5325-5334, doi:10.1074/jbc.M116.730523 (2017).

49 Stevens, G. *et al.* Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. *Eur J Public Health* **23**, 146-152, doi:10.1093/eurpub/ckr176 (2013).

50 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).

51 Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* **49**, 834-841, doi:10.1038/ng.3841 (2017).

52 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

53 Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825, doi:10.1038/s41467-018-03621-1 (2018).

54 Saw, J. *et al.* Chromosome 1q21.2 and additional loci influence risk of spontaneous coronary artery dissection and myocardial infarction. *Nat Commun* **11**, 4432, doi:10.1038/s41467-020-17558-x (2020).

55 Akcimen, F. *et al.* Transcriptome-wide association study for restless legs syndrome identifies new susceptibility genes. *Commun Biol* **3**, 373, doi:10.1038/s42003-020-1105-z (2020).

56 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).

57 Al-Khelaifi, F. *et al.* Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. *Sci Rep* **9**, 19889, doi:10.1038/s41598-019-56496-7 (2019).

Tables

Table 1 Summary of the cohorts used in this study.

Categories	Chinese cohort		European cohort	
	Cases (n = 89)	Controls (n = 209)	Cases (n = 578)	Controls (n = 583)
Age, years old				
Range	21-26	20-26	55-65	55-65
Mean (SD)	23.8 (1.6)	23.3 (1.6)	61.2 (3.1)	61.4 (3.1)
Sex, n (%)				
Male	89 (100)	209 (100)	286 (49.48)	300 (51.46)
Female	0 (0)	0 (0)	292 (51.52)	283 (48.54)

SD, standard deviation.

Table 2 Association results of the four significantly associated genes in the Chinese cohort and the European cohort.

No.	Chr.	Genes	Chinese cohort (n = 298)			European cohort (n = 1,161)			Meta-analyses (n = 1,459)		
			Beta	SE	*P*	Beta	SE	*P*	Beta	SE	*P*
1	2q23.3	*ARL6IP6*	27.57	13.33	3.95×10^{-2}	94.9	15.89	3.09×10^{-9}	55.38	10.21	5.85×10^{-8}
2	2q11.2	*TMEM127*	-0.98	0.49	4.85×10^{-2}	-2.24	0.47	2.69×10^{-6}	-1.64	0.34	1.78×10^{-6}
3	7p15.3	*TOMM7*	0.25	0.12	4.53×10^{-2}	0.57	0.12	4.87×10^{-6}	0.41	0.09	2.87×10^{-6}
4	11q25	*JAM3*	1.57	0.57	6.12×10^{-3}	1.91	0.51	1.73×10^{-4}	1.76	0.38	3.33×10^{-6}

SE, standard error. Chr, chromosome.

Meta-analyses were performed to combine the association results in the two cohorts. However, only 4 of the 46 genes (*ARL6IP6*, *TMEM127*, *TOMM7* and *JAM3*) showed nominal significance at $P < 0.05$ in both cohorts, and the association directions for these 4 genes were consistent across the two cohorts.

Figures

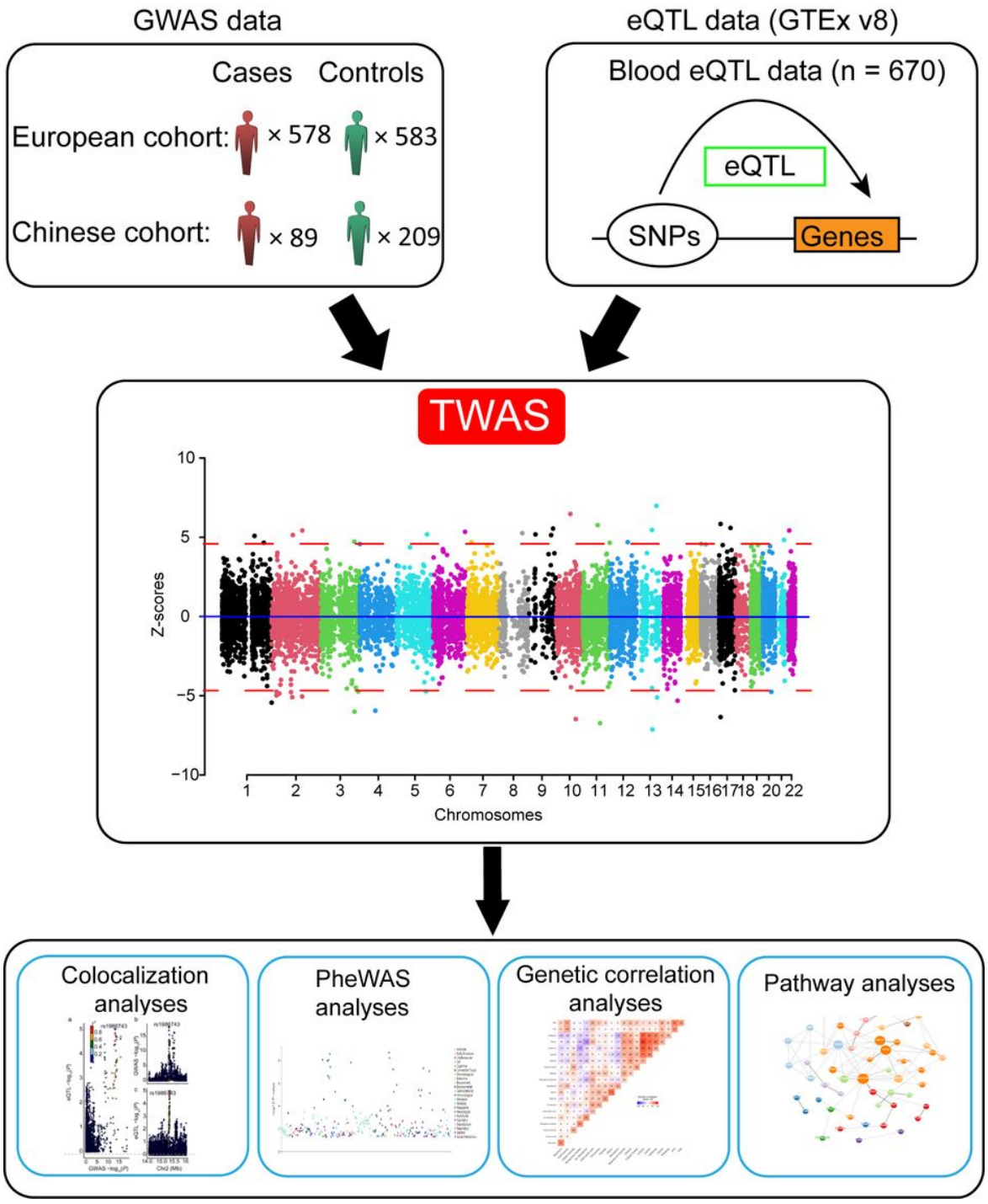

Figure 1

Schematic workflow of this study. We performed a transcriptome-wide association study (TWAS) for the hearing impairment (HI) based on the publicly available genome-wide association study (GWAS) datasets and expression quantitative trait loci (eQTL) dataset. The GWAS datasets were from a Chinese cohort (n = 298) and an European cohort (n = 1,161). The eQTL dataset was from 670 blood tissues from Genotype-Tissue Expression (GTEx) (v8). Follow-up analyses, including the colocalization analyses, phenome-wide association analyse (pheWAS), genetic correlation analyses and pathway analyses, were performed to extensively characterize the identified associations. SNP, single nucleotide polymorphism.

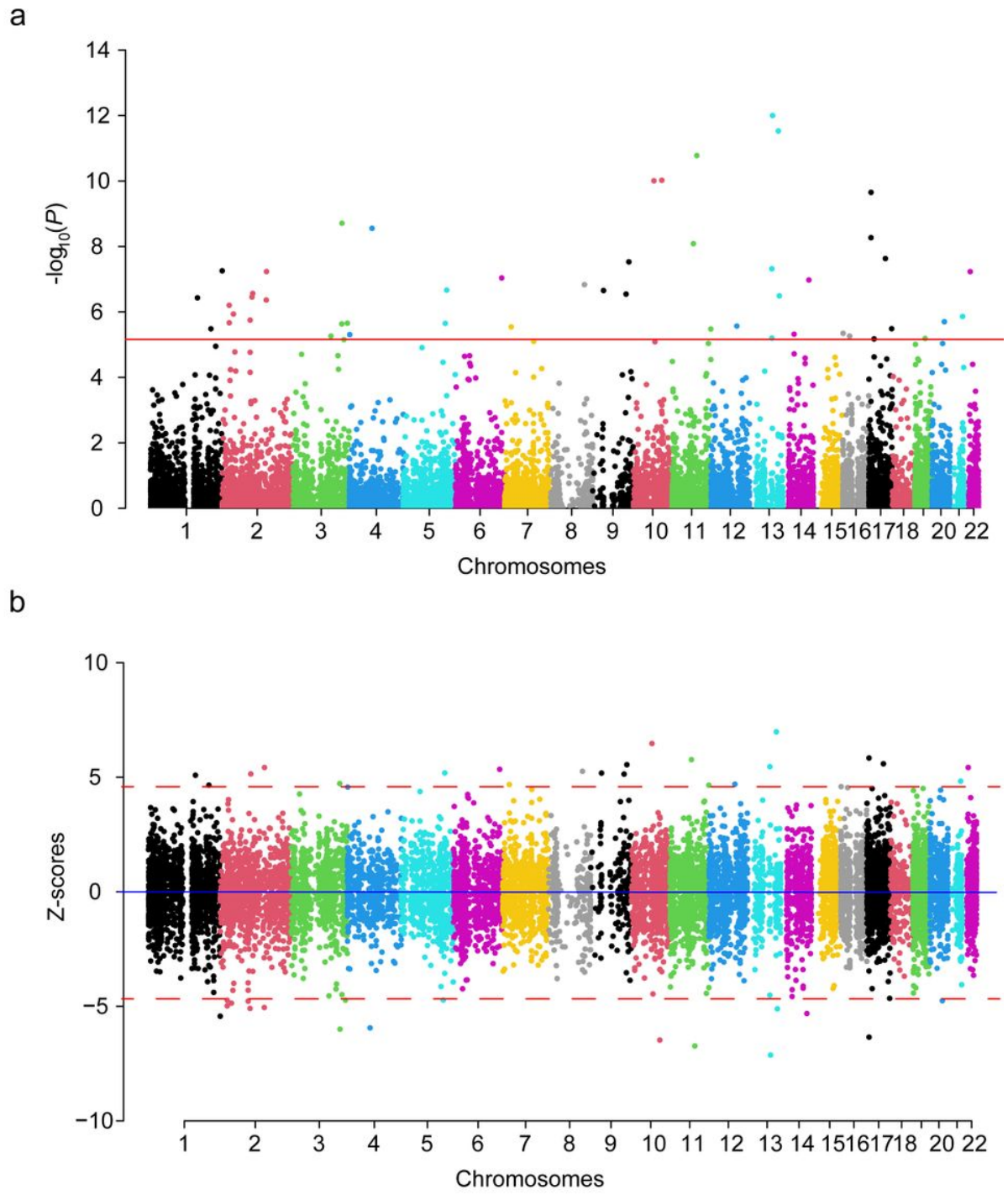

Figure 2

Manhattan plots of the TWAS results. (a) Gene-level Manhattan plot of the meta-analyses of the TWAS results from the European cohort and Chinese cohort. The x-axis represents the genomic position (based on NCBI Build 37), and the y-axis shows the -log10 (P). The red line represents the transcriptome-wide significance threshold (P = 3.96 × 10-6). (b) Z-scores of the meta-analyses of the TWAS results from the European cohort and the Chinese cohort. The x-axis represents the genomic position (based on NCBI Build 37), and the y-axis shows the Z-score from the association tests. The blue line indicates that Z-score is equal to 0. Red dotted

lines denote the Bonferroni-corrected significance threshold (|Z| = 4.65, PTWAS < 3.96 × 10-6). TWAS, transcriptome-wide association study.

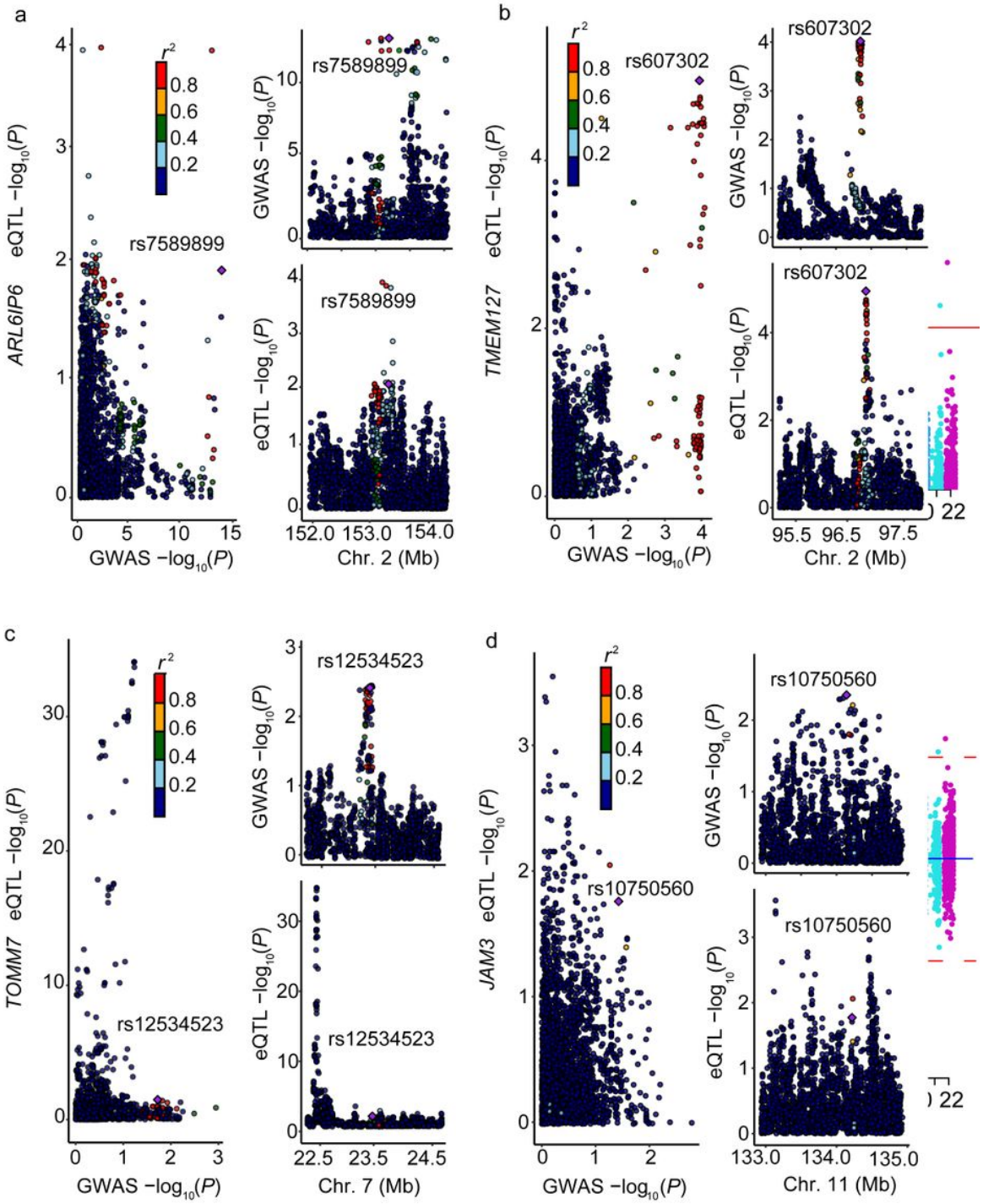

Figure 3

The locus-compare scatter plot for the association signals at ARL6IP6, TMEM127, TOMM7 and JAM3 in the European cohort. Colocalization analyses results are shown for ARL6IP6 (a), TMEM127 (b), TOMM7 (c), and JAM3 (d) loci. The locus-compare scatter plot compares the expression quantitative trait loci (eQTL) results and the genome-wide association study (GWAS) results, which indicates whether the GWAS top locus is also the leading SNP in the eQTL result. The eQTL results were based on 670 blood tissues from Genotype-Tissue

Expression (GTEx) (v8). The GWAS results were from the European cohort (n = 1,161). The gene prioritized in each locus is shown on the y-axis of the corresponding figure label. Chr., chromosome.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx