# GWAS of Chronic Dizziness in the Elderly Identifies Significant Loci Implicating *MLLT10, BPTF, LINC01225*, and *ROS1*

Royce Clifford*,[1,2] Daniel Munro,[3,4] Daniel Dochtermann,[5] Poornima Devineni,[5] Saiju Pyarajan,[5] Million Veteran Program,[5] Francesca Talese,[3] Abraham A. Palmer,[3,6] Pejman Mohammadi,[4] and Rick Friedman[1]

[1]Department of Otolaryngology-Head and Neck Surgery. University of California San Diego, La Jolla, CA, 92037, USA

[2]Veteran Administration Hospitals, San Diego, California, 92161, USA

[3]Department of Psychiatry, University of California San Diego, La Jolla, CA, 92037, USA

[4]Department of Integrative Structural and Computational Biology, Scripps Research, La Jolla, 92037, CA, USA

[5]Million Veteran Program, Veterans Administrations Hospitals, Boston, MA, 02130, USA

[6]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

*Correspondence: r2clifford@health.ucsd.edu

**Background**: Chronic age-related dizziness arises from dysfunction of the balance system, an elegant neuroanatomical group of pathways that mediates our perception of linear acceleration, gravity, and angular head motion. Studies indicates that 27-46% of chronic imbalance is genetically inherited, nevertheless, underlying genes leading to chronic imbalance remain unknown. **Subjects and Methods:** The population comprised diverse-ancestry participants in the Million Veteran Program. Cases consisted of two diagnoses of dizziness at least six months apart, excluding acute vertiginous syndromes, ataxias, syncope, and traumatic brain injury. Genome-wide association studies were performed as separate logistic regressions on Europeans, Europeans of age ≥ 50, African Americans, and those of Hispanic ancestry, followed by trans-ancestry meta-analysis. Downstream analysis included case-case-GWAS, fine-mapping, probabilistic colocalization of significant variants and genes with eQTLs, and functional analysis of significant hits. **Results:** The final cohort consisted of 50,339 cases and 366,900 controls. Two significant loci were identified in Europeans, another in the Hispanic population, and two additional loci in trans-ancestry meta-analysis. Fine mapping revealed credible sets of intronic single nucleotide polymorphisms in genes including *MLLT10* - a histone methyl transferase cofactor, *BPTF* - a subunit of a nucleosome remodeling complex implicated in neurodevelopment, *LINC01225* - affecting transcription of *ZNF91*, a repressor of retrotransposons, and *ROS1* – a proto-oncogene receptor tyrosine kinase. **Discussion:** Balance dysfunction can lead to catastrophic outcomes, including falls, injury, and death in the elderly. By honing the phenotype to be appropriate for chronic age-related dizziness, findings suggest genomic candidates for further study and ultimate treatment of this common neurologic disease.

## Introduction

Dizziness arises from dysfunction of the balance system, requiring an integration of sensory input in large part from the vestibular system. This complex cochleovestibular organ is an elegant and highly conserved neuroanatomical group of pathways that mediates our ability to perceive linear acceleration, gravity, and angular head motion. Amalgamation of vestibular, visual, and proprioception input occurs in the cerebellum, with further processing in the cerebral cortex. Disorders of the vestibular system are a major source of falls,[1] a leading cause of death in the elderly. Epidemiological evidence suggests that over 6 million individuals in the US experience chronic disequilibrium that can lead to falls.[2] The incidence of dizziness increases from 22% for adults between 65 and 69 years of age to over 40% for adults between the ages of 80 and 84 years [3–5] and there is an abundance of literature concerning associated comorbidities related to dizziness in the elderly, including increased falls, depression, social isolation, fear, and overall functional decline.[4,6] Thus, while falls in the elderly can have multiple causes, dizziness is a strong predictor of morbidity and mortality.

In younger patients, dizziness presents acutely as vertigo or a whirling sensation emanating from stimuli from the three semicircular canals, utricle, and saccule in the inner ear. In contrast, as the vestibular system undergoes age-related degeneration,[7,8] complaints become vague, and are more often described as an instability, dizziness, or an insecure gait, making diagnosis, treatment, and phenotyping more difficult.[9] The definition of "dizziness" itself is a broader term than "vertigo", nevertheless, older patients report more frequently report unsteadiness and swaying than a spinning sensation to the healthcare provider.[9] Although the recently coined term "presbyvestibulopathy" has stricter diagnostic criteria, including objective clinical testing, and is uncommon in patients hospitalized with complaints of vertigo and

moderate impairment,[10] here we use a predominantly ambulatory cohort to identify those with the health-record diagnosis of chronic dizziness.

Although dizziness in the elderly encompasses multiple systems and connections, here our aim is to focus on the inner ear's role in chronic imbalance by carefully defining inclusion and exclusion criteria in the electronic health record (EHR). Animal studies document significant age-related degeneration in nearly all types of vestibular cells, including the sensory end organ hair cells, nerve fibers, and Scarpa ganglion cells.[7,11–16]

Evidence of a genetic component to balance dysfunction in the elderly includes a large Swedish population-based twin study, where self-estimated impaired balance was assessed as a risk for osteoporotic fractures and demonstrated an age- and sex-adjusted heritability of 27%.[17] An Australian twin study, using the Lord's Balance test and Step test, found heritability of 46% for sensory balance modules and 30% for static and dynamic perturbations.[18]

"Minimal phenotyping" is defined as dependence on self-report, i.e., subjective "symptoms" rather than objective "signs" for case definitions and can degrade the quality and specificity of GWAS results.[19] It has been shown that even studies that use a single diagnosis in the electronic health record, implying an examination by a healthcare provider, can elicit a low sensitivity of 50% and specificity of 81%, leading to high rates of false positives in subsequent genome-wide association studies (GWAS), despite single examination by a clinician.[19,20] We have attempted to address this problem of self-report and individual examinations using the richness of the data in the Million Veteran Program (MVP).[21] Although "dizziness" or the report of a whirling sensation is a symptom rather than a finding, we use International Classification of Diseases (ICD) codes, since a health provider had performed a clinical examination on at least

two occasions six months apart or more and that the imbalance had continued beyond an acute episode.

In an attempt to narrow the physiologic definition, the phenotyping was designed to remove acute vestibular diseases, as well as primary cerebellar and other intracranial diseases of imbalance. Although a large portion of chronic dizziness in the elderly includes the diagnosis of benign paroxysmal positional vertigo (BPPV),[9,10] this was excluded, as the primary focus was the aging cochlea. Other removed diagnoses were acute and episodic vestibular syndromes such as Meniere's Disease, vertigo associated with migraine, and cerebellar ataxias including Parkinson's Disease.

Here we report GWAS with analyses of Europeans (EU), Americans of African descent (AA) and Hispanic ancestries (LA), then combined ancestries in meta-analysis. Downstream analysis seeks to define genes and single nucleotide polymorphisms (SNPs) in regard to colocalization with expression quantitative trait loci (eQTLs) in specific tissues, case-case GWAS to identify different allele frequencies specific to age, and fine-mapping suggesting plausible causal variants.

## Subjects and Methods

*Participants and Phenotyping*

Of the 819,397 subjects in MVP v_20_1 release, 648,224 had been assigned ancestry according to a harmonized ancestry and race/ethnicity genomic algorithm (Figure 1).[22] Chronic dizziness was assessed as a binary variable with a case defined as two diagnoses of dizziness or vertigo at least six months apart in the VA EHR. Excluded cases were those not genotyped, with no indication of sex, of only one diagnosis of dizziness or vertigo, those with diagnoses less than 6 months apart, ICD diagnosis of an acute vertiginous syndrome such as BPPV and vestibular neuronitis. Causes of the presenting complaint of dizziness not directly related to the cochlea, such as syncope, ataxia including Parkinson's Disease, and any evidence of prior traumatic brain injury (TBI) were excluded as well (See Supplementary Material Methods for list of ICD codes and numbers of participants excluded and Table e1). Since there were < 10,000 East and South Asian ancestry participants, we did not include them because the sample size was too small to be statistically informative. Participants assigned to the control group carried no diagnoses related to dizziness or vertigo.

The final sample consisted of 417,239 participants including 50,339 cases and 366,900 controls in three ancestries. Participants had a mean age of 61.9 ± 14.3. The cohort consisted of 303,743 of European ancestry (72.80%); 79,388 participants of African American descent (19.02%); and 34,108 with Hispanic ancestry (8.17%), which is broadly consistent with the US military population. Males constituted 379,968 (91.06%), and 37,271 (8.93%) were female, also consistent with the make-up of the US military. Of the total, 255,559 were ≥ 50 years of age (61.25% of the cohort).

Since there was no applicable replication cohort, we compared our phenotype of chronic dizziness to a previously published meta-analysis of a European cohort of 48,072 cases and 894,541 controls with any diagnosis of vertigo.[23] Although this cannot be construed as a "replication" cohort, to our knowledge there were no chronic dizziness GWAS studies available in the literature for direct comparison. Nevertheless, the previously published study provided a comparison of acute versus chronic dizziness of cochlear origin, since ours was designed to exclude other-than-cochleovestibular sources of dizziness.

*Genetic and Statistical Analysis*

Genotyping had been carried out by the MVP Core via a 723,305–SNV biobank array (Axiom; Affymetrix), customized for MVP to include variants of interest in multiple diverse ancestries.[24] Imputation had been performed with Minimac 3 and the 1000 Genome Project phase 3 reference panel. Analyses used MVP Release 4 data (GrCh37). Final genotype data consisted of 96 million genetic variants. For each ancestry group, principal components of genotype were calculated using PLINK 2.0 alpha.[25] See Supplemental Content for method details.

GWAS was performed as a logistic regression testing imputed dosages for association with chronic balance under an additive model using REGENIE v2.2.,[26] including the first 10 principal components of genotype, sex, and age as covariates. REGENIE is a two-step machine learning method that accounts for relatedness. Step One consisted of analysis of MVP genotype array data only, where SNPs are partitioned into consecutive blocks, with ridge regression predictions generated from each block. Predictions are then combined in a second ridge

regression procedure and decomposed into separate chromosomes for leave-one-chromosome-out (LOCO) analysis. These LOCO predictions then became covariates in Step Two. In Step Two, Release 4 imputed data was used for cross validation. Firth logistic regression and saddle point approximation was used for this binary trait.

GWAS was analyzed separately on each of the three ancestries (EU, AA, and LA), with 913,319 SNPs filtered at MAF > 0.01, INFO > 0.60, and HWE > 1 x $10^{-03}$. Initially, GWAS was calculated for two related phenotypes: two diagnoses ≥ 90 days apart on each ancestry - EU, AA, and LA ancestries (N = 427,959, cases = 55,404, controls = 372,555) and ≥ 180 days apart (N = 417,230, cases = 50,339, controls = 366,00). Since sample sizes including participants ≥ 50 years of age were sufficient only for EU, we then performed GWAS separately on this subset.

Since results were not substantially different ($r_g$ = 1.00, P-value 0.0, see Table e2), here we report the results from the ≥ 180 days diagnosis only. Summary statistics were filtered to minor allele frequency of > 1% and imputation information score > 0.6. For primary analyses, genome-wide significance was defined as $P < 5.0 \times 10^{-08}$. Regional association plots were generated using LocusZoom with 400-kilobase windows around the index variant and linkage disequilibrium patterns calculated based on 1000 Genomes Project ancestries appropriate for each lead SNP.[27]

After GWAS on individual ancestries, summary statistics were combined in meta-analysis using METAL.[28] Results of each study were weighted proportional to the square-root of each study's sample size. 17,735,101 variants remained after filtering with INFO > 0.6, MAF ≥ 0.01, and inconsistent allele labels and strands adjusted or removed. Cochran's *Q*-test was implemented for each SNP.

We calculated $h^2$ values on GWAS summary data of MVP European participants (N = 303,743) with linkage disequilibrium score regression (LDSC). [29] QQ Plots for each GWAS are in Figure e1. The proportion of inflation of test statistics due to the polygenic signal (rather than population stratification) was calculated with linkage disequilibrium score regression (LDSC), specifically, as 1 − (LDSC intercept −1)/(mean observed $\chi 2-$ 1).[29]

For comparison, we used a publicly available GWAS summary statistic dataset (n = 48,072 and 894,541 controls from Iceland/Finland/UKB/US) for vertigo, which similarly focuses on the vestibular system.[23] Replication was considered significant at P-value < 0.00625, a Bonferroni adjustment for the eight significant lead SNPs from our GWAS. The Ice/Fin/UKB/US phenotype had been defined as a single EHR diagnosis of vertigo, including ICD9 386.x and ICD10 of H81.xx., which includes acute syndromes.[23] GWAS summary stats from each cohort had been combined in a fixed-effects inverse variance model.

The proportion of heritability explained by our GWAS was performed with linkage disequilibrium scores (LDSC).[30] Pairwise genetic correlations ($r_g$) between results on chronic imbalance, hearing,[31] and tinnitus[32] were calculated using bivariate LDSC regression, with publicly available GWAS summary statistics. LDSC regressions were computed with the Ice/Fin/UKB/US replication cohort for comparison.

*Post-GWAS analysis*

Case-case GWAS compared lead SNPs identified in this study to the replication cohort, using CC-GWAS in R in order to test for differences and directions of effect in allele frequencies between the two (see Supplementary Methods).[33] Input consisted of effect sizes,

effective N's for cases and controls, P-values, population statistics, heritability for the disorders, and overlap between the two cohorts. For significant SNPs (P-value = $5.0 \times 10^{-08}$), output included statistically significant effect sizes reflecting direction and magnitude of effect and measured genetic distance between the two cohorts.[33]

Next, probabilistic colocalization between the GWAS associations (EU ancestry only) and expression quantitative trait loci analysis (eQTLs) was performed using fastENLOC to identify genes whose expression may mediate the GWAS association.[34,35] We used the pre-computed eQTL annotations for GTEx v8 data from the fastENLOC repository, with SNPs labeled with rsIDs.[36] Approximately independent LD blocks for European ancestry from LDetect were also included as input.[37] fastENLOC was run separately on the eQTL data from each of the 49 tissues, and regional colocalization probability (RCP) per eQTL per tissue were examined for evidence of colocalization. We used the RCP > 0.1 cutoff for putative causal gene contributors from the original fastENLOC study.

Fine mapping of European GWAS hits was computed with a "Sum of Single Effects" for summary statistics model using default settings in R.[38] susieR v0.12.27 is a Bayesian method to assess uncertainty in which variables are selected, involving highly correlated variables. susieR provides "Credible Sets" of variables which include relevant SNPs and a posterior inclusion probability (PIP) for each variable based on effect sizes and linkage disequilibrium (LD) between SNPs. Minimum LD is 0.5 within each CS, and indicates that within each CS, one of the variables has a non-zero effect coefficient. PIP was set at 0.95. See details in Supplemental methods. Originally formulated for linear regression, susieR has been verified for logistic regression.

Functional mapping and annotation of genetic associations (FUMA), version 1.4.2, was examined (Figures e2-e9),[39] based on human genome assembly GRCh37 (hg19). FUMA was used with default settings unless stated otherwise. Gene-based analysis was performed with the FUMA implementation of MAGMA version 1.08.[40] Significance of genes was determined with a Bonferroni-corrected threshold of P-value = 0.05/18,914 protein coding genes = $2.644 \times 10^{-06}$. eQTL gene mapping was included using GTEx v8 Brain with 13 tissues. 3D chromatin interaction mapping was performed to capture chromatin conformation with dorsolateral prefrontal cortex and hippocampus tissues (HiC GSE87112).[41]

PheWAS was performed using SNPs with the smallest P-value from fine-mapping at the Atlas of GWAS Summary, with significance based on Bonferroni correction for the number of phenotypes engaged by each SNP.[42]

*Ethics and Inclusion Statement*

This research has included local researchers throughout the research process, as is evidenced by authorship, including study design, implementation, data ownership, intellectual property, and authorship of publications. This research is locally relevant to the Veteran Administration population. Roles and responsibilities were agreed upon ahead of the research. This research would not have been severely restricted or prohibited in the setting of the local researchers.

Prior written informed consent for use of health records and genomic data was obtained for all participants in this retrospective analysis within the Million Veteran Program. This study was approved by Central VA IRB reference E22-36, qualifying for exemption under category

4(ii). Researchers in this study have access to only de-identified individual data as protected by Veteran Administration protocol, thus there is no access to individuals and no identification of any of the subjects. As described in IRB approval, risk management plans ensure privacy and protection of individuals engaged in the study. Since this is a data-only study, no biological materials, cultural artifacts, or associated traditional knowledge has been transferred out of the country.

## Results

*Demographics*

Chronic dizziness was associated with increased age, P-value < 0.001 (Table 1). Overall case prevalence was 12.06%, with case mean age 67.21 ± 11.67 versus control mean age 61.15 ± 14.45, range (19 to 108), P-value < $2.0 \times 10^{-16}$. Women had a dizziness prevalence of 12.48% versus 13.67% for men (P-value < $2.0 \times 10^{-16}$). Those of African ancestry were less likely to have a EHR diagnosis of chronic dizziness than those of European or Hispanic ancestry (OR 0.728, 95% CI 0.756, 0.700, and OR 0.875, 95% CI 0.841, 0.909, respectively), P-value < $2.0 \times 10^{-16}$ and $1.97 \times 10^{-10}$.

The Ice/Fin/UKB/US cohort consisted of 48,072 cases and 894,541 controls aggregated from four European cohorts.[23] Case rate was reported as 5.10%. Although there have been other studies on motion sickness, this was the largest study identified that was available on dizziness/vertigo for comparison.[23,43]

*Heritability and Genetic Correlation of Dizziness with Other Traits*

SNP heritability for those of European descent was $h_s^{SNP}$ of 0.059 ± 0.0083 on the liability scale, assuming sample prevalence of 12% and population prevalence of 30%.[4] QQ plots identify a GCλ = 1.127, indicative of inflation of the test statistics (Fig. e1). LDSC intercept = 1.027, thus polygenic effects account for 94.87% of observed inflation.

LDSC genetic correlation compared genomic dizziness results to other cochlear disorders in EU ancestries (Table e1). Our dizziness phenotype was moderately correlated with hearing difficulties, $r_g$ = 0.314 ±0.048 (P-value $5.77 \times 10^{-11}$), and tinnitus, $r_g$ = 0.20 ± 0.8 (P-value = 1.51

x $10^{-02}$). The Ice/Fin/UKB/US showed similar correlations for hearing, $r_g = 0.19 \pm 0.049$ (P-value 2.0 x $10^{-4}$) and tinnitus, $r_g = 0.19 \pm 0.049$ (P-value 8.7 x $10^{-05}$).

*Genome-Wide Association Study and Replication*

GWAS of chronic dizziness on EU ancestry indicated two significant loci, lead hits rs10828248 on chromosome 10 (P-value 3.17 x $10^{-08}$) and rs7221651 on chromosome 17 (P-value 3.71 x $10^{-10}$) (Figure 2a). GWAS on the Hispanic ancestry indicated a novel locus on chromosome 6, rs71717606 with P-value 4.84 x $10^{-10}$, unobserved in any of the other GWAS' (Figure 2b). GWAS on the AA ancestry cohort revealed no significant findings (Figure 2c). Regional association plots are seen in Figures e2-e9.

GWAS in trans-ancestry analysis identified four significant loci (Figure 2a and Table 2) at P-value < 5 x $10^{-08}$, including the two loci on chromosomes 10 and 17 identified in the EU GWAS (Figure 2d). Lead SNPs in chromosomes 10 and 17 from trans-ancestry and European GWAS' were in moderate to high linkage disequilibrium (LD). $R^2$ of rs12779865/rs10828248 on chromosome 10 = 0.87 and $R^2$ of rs72631329/rs7221651 on chromosome 17 = 0.54.

Comparison of GWAS on all Europeans to those ≥ 50 years of age yielded similar results, with no significant changes in z-scores of the identified loci on chr 10 and 17 (data not shown, Cohen's D = 0.001 for both hits, indicative of essentially no difference in the two cohorts). Once again, lead SNPs for EU ≥ 50 were in LD with both GWAS of all EU participants and with trans-ancestry GWAS results, with an identical lead SNP on chr10. On chromosome 17, rs7221651/rs4791046 $R^2 = 0.99$ for EU versus EU ≥ 50 years of age.

The MVP chronic dizziness phenotype and Ice/Fin/UKB/US vertigo phenotype demonstrated a genetic correlation of $r_g = 0.67 \pm 0.073$, P-value $5.34 \times 10^{-20}$ (Table e1). Despite differences in phenotypes, two loci were replicated in the Ice/Fin/UKB/US cohort (Table 2), lead SNPs on chromosome 10, rs12779865/rs10828248 (P-values $3.98 \times 10^{-03}$ and $4.79 \times 10^{-03}$, respectively) and chromosome 19 rs10713223 (P-value $1.45 \times 10^{-27}$).

Gene set analysis identified seven mapped genes at P-value < 0.05/19,093 = $2.619 \times 10^{-06}$ for 19,093 protein coding genes analyzed (Table 3). For the gene set analysis, two of the several genes identified in our cohort were replicated: *ZNF19* (P-value $6.34 \times 10^{-14}$) and *MLLT10* (P-value $1.76 \times 10^{-03}$).[40]

Case-case GWAS (CC-GWAS) is a method to test for genetic differences between allele frequencies of two disorders.[33] Comparison with Ice/Fin/UKB/US indicated a sole significant difference in alleles on the chromosome 19 locus (Table e3), with lead SNP rs17473980 (P-value $1.52 \times 10^{-23}$).

We tested for colocalization between the European GWAS associations and eQTLs in 49 tissues assayed by the GTEx Project.[36] Regional colocalization probabilities were below 0.1 for all eQTLs, indicating lack of evidence for gene expression in the tested tissues mediating the genetic associations with chronic dizziness. The strongest colocalizations were with *MIR1915HG* (*CASC10*) in Aorta (RCP = 0.068) and *MLLT10* in sun-exposed skin (RCP = 0.064), while all other eQTLs had RCP < 0.03.

Fine mapping was performed within significant loci using summary statistics and an LD matrix from 1000 genomes with ancestry-matched values.[38] Table 4 and Figure e10 show a list of "credible sets" with a posterior probability > 0.95 for loci identified as significant, and their

linkage disequilibrium $r_g$ with the lead SNP in each locus. For the chromosome 6 sets, fine mapping identified SNPs within an intron of *ROS1*. The sets within chromosome 10 are within an intron of *MLLT10* and in high LD with the lead SNP. In chromosome 17, the five identified SNPs are in equilibrium with the lead SNP, thus no conclusion as to "cause" can be made, however, they are all within introns of *BPTF*. In chromosome 19, both the lead SNP and the two identified SNPs are within a long non-coding RNA, *LINC01224*, in contrast to the Ice/Fin/UKB/US GWAS, where the lead SNP was within *ZNF19*. This finding is in agreement with CC-GWAS results, which indicated a significant difference in alleles within this locus.

*Integration with Functional Genomic Data*

We examined 54 body tissues for gene enrichment using sumstats from the EU GWAS and found that although ataxias and Parkinson's disorders were excluded from the phenotype, the cerebellum was nevertheless significantly enriched at P-value $9.4 \times 10^{-04}$ (Figure 3 and Table e4). When reanalyzed with 3D chromatin interaction mapping based on Hi-C and Chia-PET, frontal cortex and brain cortex genes became significantly overexpressed.[39,44] We then looked at functional analysis of lead variants for all ancestries (Table e5 and Figures e2-9). All significant loci contained multiple SNPs with combined annotation dependent depletion (CADD) scores > 12.37 and RegulomeDB scores < 5, indicating that the SNP alteration is within the top 7% of deleterious substitutions and the locus contains a transcription factor binding site, a Dnase peak, and/or eQTL.[45] It should be noted that these tissues do not include the cochleovestibular system, however.

PheWAS of > 3,300 traits and disorders for each significant SNP in the fine mapping analysis showed multiple correlations (Table e6 and Figure e11 for lead SNPs). From the chr6 rs72965321 locus, the most significant hit was a correlation with phenotypes for skin color (P-value = $1.87 \times 10^{-05}$) and hippocampus anisotropy in diffusion tract imaging (DTI), indicating loss of white matter tracts in the hippocampus (P-value = $8.44 \times 10^{-05}$). The chr10 rs8904565 hit was significant for multiple areas of the brain, including anisotropy of the fornix (P-value = $3.39 \times 10^{-11}$), posterior limb of the internal capsule (P-value $4.87 \times 10^{-09}$), left accumbens, total brain volume, posterior corona radiata, and intra-cranial areas. Chr17 rs7221651 was significant for multiple impedance measures and other measurements of body mass (Impedance measurement of trunk fat mass, P-value $7.05 \times 10^{-30}$. Chr19 rs612285 showed minimal significance for frequent insomnia (P-value $1.40 \times 10^{-04}$) and menopause (P-value $2.89 \times 10^{-04}$).

**Discussion**

Good balance requires integration of sensory input from cochlea, proprioception, and the eye, with translation to motor output to the eye and body muscles. Vestibular dysfunction, as measured by a modified Romberg, increases with age, and is measured in 35.4% of US adults over 40 years. Imbalance carries a 12x increase in the odds of falling, leading to morbidity and mortality among older individuals.[46] Part of the reason for patients' lack of report of vertigo is related to central compensation of visual and proprioceptive, and these systems degenerate with age as well.[4]

The non-specific report of dizziness can lead to a myriad of diagnoses, depending on testing and findings, including benign paroxysmal positional vertigo (BPPV), unilateral

vestibulopathy, orthostatic hypotension, low vision, proprioceptive impairment, and extrapyramidal disorders, among others.[47] In order to focus on sensory input in the aging cochlea, we avoided "minimal phenotyping", defined as dependence on one self-report for case definitions.[19] For chronicity, we sought participants who had at least two clinical diagnoses six months apart. Our reasoning was that studies that use a single diagnosis in the electronic health record can elicit a low sensitivity of 50% and specificity of 81%, leading to high rates of false positives in subsequent GWAS, despite examination by a clinician.[19,20] In order to focus further on chronic dizziness, we excluded acute and intermittent vertiginous syndromes (i.e., Meniere's Disease, benign paroxysmal positional vertigo, and acute vestibular neuronitis), as well as non-vestibular system gait disorders such as Parkinson's and other ataxias. We also eliminated over 25,000 of those diagnosed with traumatic brain injury (TBI), since TBI is a strong predictor of vertigo that is presumed to have a separate etiology.[48]

Our study shows that other cochlear disorders are genetically correlated to this GWAS finding, consistent with previous studies. While a replication cohort with the same phenotype does not exist to our knowledge, and despite differences in phenotype definition, genetic correlation with Ice/Fin/UKB/US was still reasonably strong and highly significant ($r_g$ = 0.67 se 0.07, P-value 5.34 x $10^{-20}$), indicating that the "chronic dizziness" phenotype shared common genetic influences with other acute vestibular syndromes.

In post-GWAS analysis, we elicited potential causal SNPs and genes from significant loci. All lead SNPs and significant fine-mapping SNPs identified were intronic or within a long non-coding RNA. SNPs identified by fine mapping include intronic variants in the gene histone lysine methyltransferase *DOT1L* cofactor (*MLLT10*) within an enhancer region.[49] *MLLT10* is expressed in the vestibule and the cochlea[50] and is thought to prevent somatic cell

reprogramming through regulation of *DOT1L*-mediated H3K79 methylation. A SNP in LD with the top hit rs10828248 has been associated with the genesis of meningiomas ($R^2$ = 0.8403, P-value < 0.001).[51]

Fine mapping identified a credible set of intronic SNPs in the gene bromodomain PHD finger transcription factor (*BPTF*) in chromosome 17. *BPTF* is the largest subunit of the nucleosome remodeling complex NURF and has been implicated in neurodevelopmental disorders, developmental delay, and autism.[52] While pheWAS has identified predominant associations with body mass rather than a neurological function, this gene is identified in Type I and Type II hair cells of the mature mouse utricle in the vestibular system in higher abundance than in the cochlea.[50]

Trans-ancestry GWAS identified rs10713223 (P-value 3.11 x $10^{-08}$), an insertion/deletion variant on chromosome 19, which replicated the locus from Ice/Fin/UKB/US study as is thus associated with vestibulocochlear disorders; however, they noted the gene of interest as the nearby *ZNF91*.[23] CC-GWAS for this locus was highly significant at P-value = 4.55x $10^{-27}$, indicating a difference in allele frequencies between the two studies. Our lead SNP and fine-mapping analysis identified a credible set of SNPs within *LINC01224*.[53] *LINC01224* affects the transcription of the nearby *ZNF91* by regulation of long-range interactions between the *ZNF91* enhancer and promoter.[53] *ZNF91* is a transcription factor specifically required to repress SINE-VNTR-Alu retrotransposons and, interestingly, is primate-specific.[54] The interplay of regulatory elements in this locus is yet to be determined.

A novel variant in the Hispanic population, rs71717606 (P-value 4.84 x $10^{-10}$) is an insertion/deletion within an intron of the gene *Discoidin, CUB,* and *LCCL* domain-containing

protein (*DCBLD1*) on chromosome 6. *DCBLD1* is a transmembrane receptor involved in development.[55] The top hit rs71717606 is located in a promoter site.[49] In contrast, fine mapping pointed to rs72965321, which is in an intron of an adjacent gene *ROS1*. eQTL analysis in FUMA reveals significant expression of *ROS1* in the frontal cortex, with significant HiC loops in the hippocampus (P-value < $2.3 \times 10^{-11}$ after Bonferroni correction),[39] indicative of enhancer-promoter and promoter-promoter interactions within *ROS1*. *ROS1* is expressed solely in the vestibular ganglion and has not been identified within the vestibular portion of the cochlea.[50,56]

Although our phenotype was crafted to eliminate causes of chronic dizziness outside the vestibular organ, nevertheless, gene enrichment identified the cerebellum as significantly enriched. This demonstrates the intimate relationship of the vestibule with the cerebellum portion of the finely tuned balance system. 3D chromatin interaction mapping added frontal cortex connections. One limitation of our study is that human vestibular expression data is not routinely included within body tissues for comparison.

This study is not without limitations. Our sample is predominantly male. The trans-ancestry meta-analysis over-sampled from Europeans and did not include individuals of Asian ancestries. The Million Veteran Program consists of a cohort that has unique battle-related environmental exposures that have poorly defined interactions with genetics and the aging process. In addition, the use of electronic health records could lead to errors of omission for several reasons. First, some of the individuals in our study received medical care outside the VA and second, patients with dizziness may perceive it to be normal for their age or be concerned with other health problems, consistent with a slightly lower prevalence of chronic dizziness identified in national surveys. Although our GWAS-eQTL colocalization analysis using GTEx

data did not identify any strong hits, those data did not include expression data from the vestibule. However, cochleovestibular animal studies indicate that these genes are all expressed within the vestibule or vestibular ganglion.[50,56] It is possible that further eQTL data from vestibular tissue or other relevant cell types could provide functional insight into the mechanisms underlying the identified genomic signal identified.

Here we have identified the first loci associated with chronic imbalance related to the cochleovestibular system, including the first significant locus in a Hispanic population. With downstream analysis including correlation of eQTL through colocalization,[34] CC-GWAS,[33] fine-mapping,[38] and associated physiologic and anatomic information indicated variants and plausible genes of interest related to lead SNPs,[57] we have identified hits containing regulatory loci, i.e., *MLLT10* and *BPTF*, a gene identified in the vestibular ganglion (*ROS1*) and a non-coding RNA, *LINC01224*, that may control its downstream gene *BPTF*, within the same locus. Ascertaining the physiologic/genetic architecture in the cochleovestibular system will aid in specific treatment of balance in the elderly. Drug development can target specifically genes related to imbalance in the elderly. Knowledge of those with significant genetic variants can be identified at higher risk for imbalance and provided more focused vestibular therapy. Future studies with larger and more gender-balanced cohorts will build on these results to clarify the role of the vestibular system in the genomics of chronic dizziness in the elderly.

This work was supported by grants from NIH R01DC020052 and Veterans Administration RR&D MVP010. Research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration. This publication does not represent the views of the Department of Veteran Affairs or the United States Government.

*Data and Code Availability*

Summary statistics of GWAS are available through the Database of Phenotypes and Genotypes (dbGaP) accession number phs001672. dbGaP requires an application requesting data per standard procedure.

Algorithms are described in the manuscript and supplemental materials. There was no custom code or mathematical algorithm formulated for this research, however, standard methods as described in other manuscripts as referenced were performed.

## References


1. Ekvall Hansson, E., and Magnusson, M. (2013). Vestibular asymmetry predicts falls among elderly patients with multi- sensory dizziness. BMC Geriatr *13*. 10.1186/1471-2318-13-77.
2. Harun, A., Semenov, Y.R., and Agrawal, Y. (2015). Vestibular Function and Activities of Daily Living: Analysis of the 1999 to 2004 National Health and Nutrition Examination Surveys. Gerontol Geriatr Med *1*, ePub. 10.1177/2333721415607124.
3. Kippenbrock, T., and Soja, M.E. (2007). Preventing falls in the elderly: interviewing patients who have fallen. Researchers attempt to identify fall-risk factors from the patients' point of view. Geriatr Nurs *14*, 205–209. 10.1016/s0197-4572(06)80043-1.
4. Fernández, L., Breinbauer, H.A., and Delano, P.H. (2015). Vertigo and Dizziness in the Elderly. Front Neurol *6*, 144. 10.3389/fneur.2015.00144.
5. Cutson, T.M. (1994). Falls in the elderly. Am Fam Physician *49*, 149–156.
6. Teixeira, A.R., Wender, M.H., Gonçalves, A.K., Freitas, C.D.L.R., Santos, A.M.P.V. dos, and Soldera, C.L.C. (2016). Dizziness, physical exercise, falls, and depression in adults and the elderly. Int Arch Otorhinolaryngol *20*, 124–131. 10.1055/s-0035-1566304.
7. Furman, J.M., and Redfern, M.S. (2001). Effect of aging on the otolith-ocular reflex. J Vestib Res *11*, 91–103.
8. tanioka, H., tanioka, S., and Kaga, K. (2020). Vestibular Aging process from 3D Physiological Imaging of the Membranous Labyrinth. 10.1038/s41598-020-66520-w.
9. Wassermann, A., Finn, S., and Axer, H. (2022). Age-Associated Characteristics of Patients With Chronic Dizziness and Vertigo. J Geriatr Psychiatry Neurol *35*, 580–585. 10.1177/08919887211036185.
10. Müller, K.J., Becker-Bense, S., Strobl, R., Grill, E., and Dieterich, M. (2022). Chronic vestibular syndromes in the elderly: Presbyvestibulopathy—an isolated clinical entity? Eur J Neurol *29*, 1825–1835. 10.1111/ene.15308.
11. Baloh, R.W., Jacobson, K.M., and Socotch, T.M. (1993). The effect of aging on visual-vestibuloocular responses. Exp Brain Res *95*, 509–516. 10.1007/BF00227144.
12. Matheson, A.J., Darlington, C.L., and Smith, P.F. (1999). Further evidence for age-related deficits in human postural function. J Vestib Res *9*, 261–264.
13. Rauch, S.D., Velazquez-Villaseñor, L., Dimitri, P.S., and Merchant, S.N. (2001). Decreasing hair cell counts in aging humans. Ann N Y Acad Sci *942*, 220–227. 10.1111/j.1749-6632.2001.tb03748.x.
14. Jang, Y.S., Hwang, C.H., Shin, J.Y., Bae, W.Y., and Kim, L.S. (2006). Age-related changes on the morphology of the otoconia. Laryngoscope *116*, 996–1001. 10.1097/01.mlg.0000217238.84401.03.
15. Walther, L.E., and Westhofen, M. (2007). Presbyvertigo-aging of otoconia and vestibular sensory cells. J Vestib Res *17*, 89–92.

16. Maes, L., Dhooge, I., D'haenens, W., Bockstael, A., Keppler, H., Philips, B., Swinnen, F., and Vinck, B.M. (2010). The effect of age on the sinusoidal harmonic acceleration test, pseudorandom rotation test, velocity step test, caloric test, and vestibular-evoked myogenic potential test. Ear Hear *31*, 84–94. 10.1097/AUD.0b013e3181b9640e.
17. Wagner, H., Melhus, H., Pedersen, N.L., and Michaëlsson, K. (2009). Heritability of impaired balance: a nationwide cohort study in twins. Osteoporosis Int *20*, 577–583. 10.1007/s00198-008-0710-3.
18. el Haber, N., Hill, K.D., Cassano, A.-M.T., Paton, L.M., Macinnis, R.J., Cui, J.S., Hopper, J.L., and Wark, J.D. (2006). Genetic and environmental influences on variation in balance performance among female twin pairs aged 21-82 years. Am J Epidemiol *164*, 246–256. 10.1093/aje/kwj188.
19. Cai, N., Revez, J.A., Adams, M.J., Andlauer, T.F.M., Breen, G., Byrne, E.M., Clarke, T.-K., Forstner, A.J., Grabe, H.J., Hamilton, S.P., et al. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet *52*, 437–447. 10.1038/s41588-020-0594-5.
20. Mitchell, A.J., Vaze, A., and Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. The Lancet *374*, 609–619. 10.1016/S0140-6736(09)60879-5.
21. Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., et al. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol *70*, 214–223. 10.1016/j.jclinepi.2015.09.016.
22. Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., Vujkovic, M., Damrauer, S.M., Pyarajan, S., Gaziano, J.M., et al. (2019). Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. The American Journal of Human Genetics *105*, 763–772. 10.1016/j.ajhg.2019.08.012.
23. Skuladottir, A.Th., Bjornsdottir, G., Nawaz, M.S., Petersen, H., Rognvaldsson, S., Moore, K.H.S., Olafsson, P.I., Magnusson, S.H., Bjornsdottir, A., Sveinsson, O.A., et al. (2021). A genome-wide meta-analysis uncovers six sequence variants conferring risk of vertigo. Commun Biol *4*, 1148. 10.1038/s42003-021-02673-2.
24. Hunter-Zinck, H., Shi, Y., Li, M., Gorman, B.R., Ji, S.G., Sun, N., Webster, T., Liem, A., Hsieh, P., Devineni, P., et al. (2020). Genotyping Array Design and Data Quality Control in the Million Veteran Program. Am J Hum Genet *106*, 535–548. 10.1016/j.ajhg.2020.03.004.
25. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience *4*, 1–16. 10.1186/s13742-015-0047-8.
26. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet *53*. 10.1038/s41588-021-00870-7.

27. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., and Frishman, D. (2011). LocusZoom: Regional visualization of genome-wide association scan results. In Bioinformatics 10.1093/bioinformatics/btq419.
28. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. 10.1093/bioinformatics/btq340.
29. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet *47*, 291–295. 10.1038/ng.3211.
30. Finucane, H, Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., ..., Neale, B., and Price, A. (2015). Partitioning heritability by functional annotation using genome- wide association summary statistics. Nat Genet *47*, 1228–1235.
31. Wells, H.R.R., Freidin, M.B., Abidin, F.N.Z., Payton, A., Dawes, P., Munro, K.J., Morton, C.C., Moore, D.R., Dawson, S.J., and Frances, M. (2019). Genome-wide association study identifies 44 independent genomic loci for self-reported adult hearing difficulty in the UK Biobank cohort. Am J Hum Genet, 1–15. http://dx.doi.org/10.1101/549071.
32. Clifford, R.E., Maihofer, A.X., Stein, M.B., Ryan, A.F., and Nievergelt, C.M. (2020). Novel Risk Loci in Tinnitus and Causal Inference With Neuropsychiatric Disorders Among Adults of European Ancestry. JAMA Otolaryngol Head Neck Surg *146*, 1015–1025. 10.1001/jamaoto.2020.2920.
33. Peyrot, W.J., and Price, A.L. (2021). Identifying loci with different allele frequencies among cases of eight psychiatric disorders using CC-GWAS. Nat Genet *53*, 445–454. 10.1038/s41588-021-00787-1.
34. Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H.K., and Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. The American Journal of Human Genetics *108*, 25–35. 10.1016/j.ajhg.2020.11.012.
35. Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y., Consortium, Gte., Wen, X., and Im, H.K. (2020). PhenomeXcan: Mapping the genome to the phenome through the transcriptome. Sci Adv *6*. 10.1126/sciadv.aba2083.
36. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979) *369*, 1318–1330. 10.1126/science.aaz1776.
37. Berisa, T., and Pickrell, J.K. (2015). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics, btv546. 10.1093/bioinformatics/btv546.
38. Zou, Y., Carbonetto, P., Wang, G., and Stephens, M. (2022). Fine-mapping from summary data with the “Sum of Single Effects” model. 10.1371/journal.pgen.1010299.

39. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nat Commun *8*, 1826. 10.1038/s41467-017-01261-5.
40. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol *11*, e1004219. 10.1371/journal.pcbi.1004219.
41. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (1979) *326*, 289–293. 10.1126/science.1181369.
42. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet *51*, 1339–1348. 10.1038/s41588-019-0481-0.
43. Hromatka, B.S., Tung, J.Y., Kiefer, A.K., Do, C.B., Hinds, D.A., and Eriksson, N. (2015). Genetic variants associated with motion sickness point to roles for inner ear development, neurological processes and glucose homeostasis. Hum Mol Genet *24*, 2700–2708. 10.1093/hmg/ddv028.
44. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep *10*, 1297–1309. 10.1016/j.celrep.2015.02.004.
45. 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*. 10.1093/nar/gky1016.
46. Agrawal, Y., Carey, J.P., della Santina, C.C., Schubert, M.C., and Minor, L.B. (2009). Disorders of balance and vestibular function in US adults: data from the National Health and Nutrition Examination Survey, 2001-2004. Arch Intern Med *169*, 938–944. 10.1001/archinternmed.2009.66.
47. Agrawal, Y., van de Berg, R., Wuyts, F., Walther, L., Magnusson, M., Oh, E., Sharpe, M., and Strupp, M. (2019). Presbyvestibulopathy: Diagnostic criteria Consensus document of the classification committee of the Bárány Society. Journal of Vestibular Research *29*, 161–170. 10.3233/VES-190672.
48. Sessoms, P.H., Gottshall, K.R., Sturdy, J., and Viirre, E. (2015). Head Stabilization Measurements as a Potential Evaluation Tool for Comparison of Persons With TBI and Vestibular Dysfunction With Healthy Controls. Mil Med *180*, 135–142. 10.7205/MILMED-D-14-00386.
49. Hunt, S.E., McLaren, W., Gil, L., Thormann, A., Schuilenburg, H., Sheppard, D., Parton, A., Armean, I.M., Trevanion, S.J., Flicek, P., et al. (2018). Ensembl variation resources. Database (Oxford). 10.1093/database/bay119.
50. Orvis, J., Gottfried, B., Kancherla, J., Adkins, R.S., Song, Y., Dror, A.A., Olley, D., Rose, K., Chrysostomou, E., Kelly, M.C., et al. (2021). gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. Nat Methods *18*, 843–844. 10.1038/s41592-021-01200-9.

51. Walsh, K.M., Zhang, C., Calvocoressi, L., Hansen, H.M., Berchuck, A., Schildkraut, J.M., Bondy, M.L., Wrensch, M., Wiemels, J.L., and Claus, E.B. (2022). Pleiotropic MLLT10 variation confers risk of meningioma and estrogen-mediated cancers. Neurooncol Adv *4*. 10.1093/noajnl/vdac044.
52. Stankiewicz, P., Khan, T.N., Szafranski, P., Slattery, L., Streff, H., Vetrini, F., Bernstein, J.A., Brown, C.W., Rosenfeld, J.A., Rednam, S., et al. (2017). Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and Dysmorphic Features. The American Journal of Human Genetics *101*, 503–515. 10.1016/j.ajhg.2017.08.014.
53. Fu, W., Zhao, J., Hu, W., Dai, L., Jiang, Z., Zhong, S., Deng, B., Huang, Y., Wu, W., and Yin, J. (2021). LINC01224/ZNF91 Promote Stem Cell-Like Properties and Drive Radioresistance in Non-Small Cell Lung Cancer. Cancer Manag Res *Volume 13*, 5671–5681. 10.2147/CMAR.S313744.
54. He, D., Wu, D., Muller, S., Wang, L., Saha, P., Ahanger, S.H., Liu, S.J., Cui, M., Hong, S.J., Jain, M., et al. (2021). miRNA-independent function of long noncoding pri-miRNA loci. Proceedings of the National Academy of Sciences *118*. 10.1073/pnas.2017562118.
55. Schmoker, A.M., Weinert, J.L., Markwood, J.M., Albretsen, K.S., Lunde, M.L., Weir, M.E., Ebert, A.M., Hinkle, K.L., and Ballif, B.A. (2020). FYN and ABL Regulate the Interaction Networks of the DCBLD Receptor Family. Molecular & Cellular Proteomics *19*, 1586–1601. 10.1074/mcp.RA120.002163.
56. Shen, J., Scheffer, D.I., Kwan, K.Y., and Corey, D.P. (2015). SHIELD: An integrative gene expression database for inner ear research. Database *2015*, 1–9. 10.1093/database/bav071.
57. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet *93*, 779–797. 10.1016/j.ajhg.2013.10.012.

**Figure Titles and Legends**

**Fig1**. Breakdown of cases and controls. Exclusion criteria and numbers in green ovals. Participants were recruited from 50 Veterans Administration Healthcare Centers across the US, beginning in 2011. Final analyzed cohort included 79,388 participants of African American descent; 303,743 of European ancestry; and 34,108 with Hispanic ancestry. Note that some excluded participants may have fit into overlapping categories.

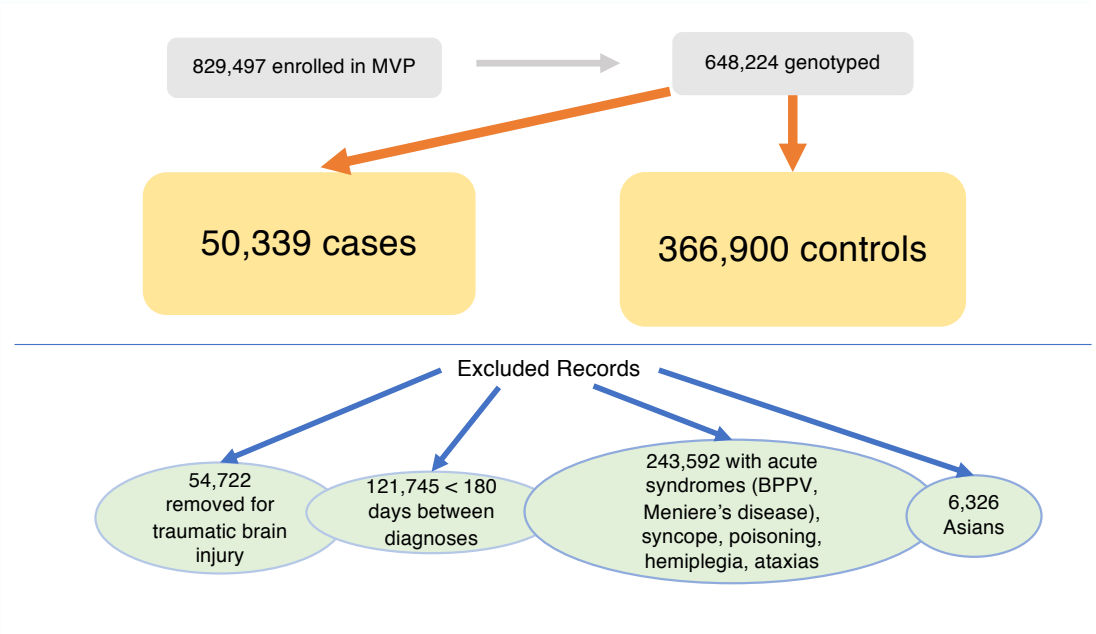


**Fig2.** Manhattan Plot of GWAS for each ancestry, with lead SNPs indicated in each locus for which P-value < 5.0 x $10^{-08}$. A. GWAS of European cohort (EU). B. GWAS of Hispanic ancestry (LA). C. GWAS of African American ancestry (AA). D. Meta-analysis of European, African American, and Hispanic populations. The horizontal line represents the Bonferroni-adjusted significance threshold P-value 5 x $10^{-08}$. Red dots indicate single-nucleotide variations within the significant regional loci (See Supplemental Figures e2-e9 for regional plots).

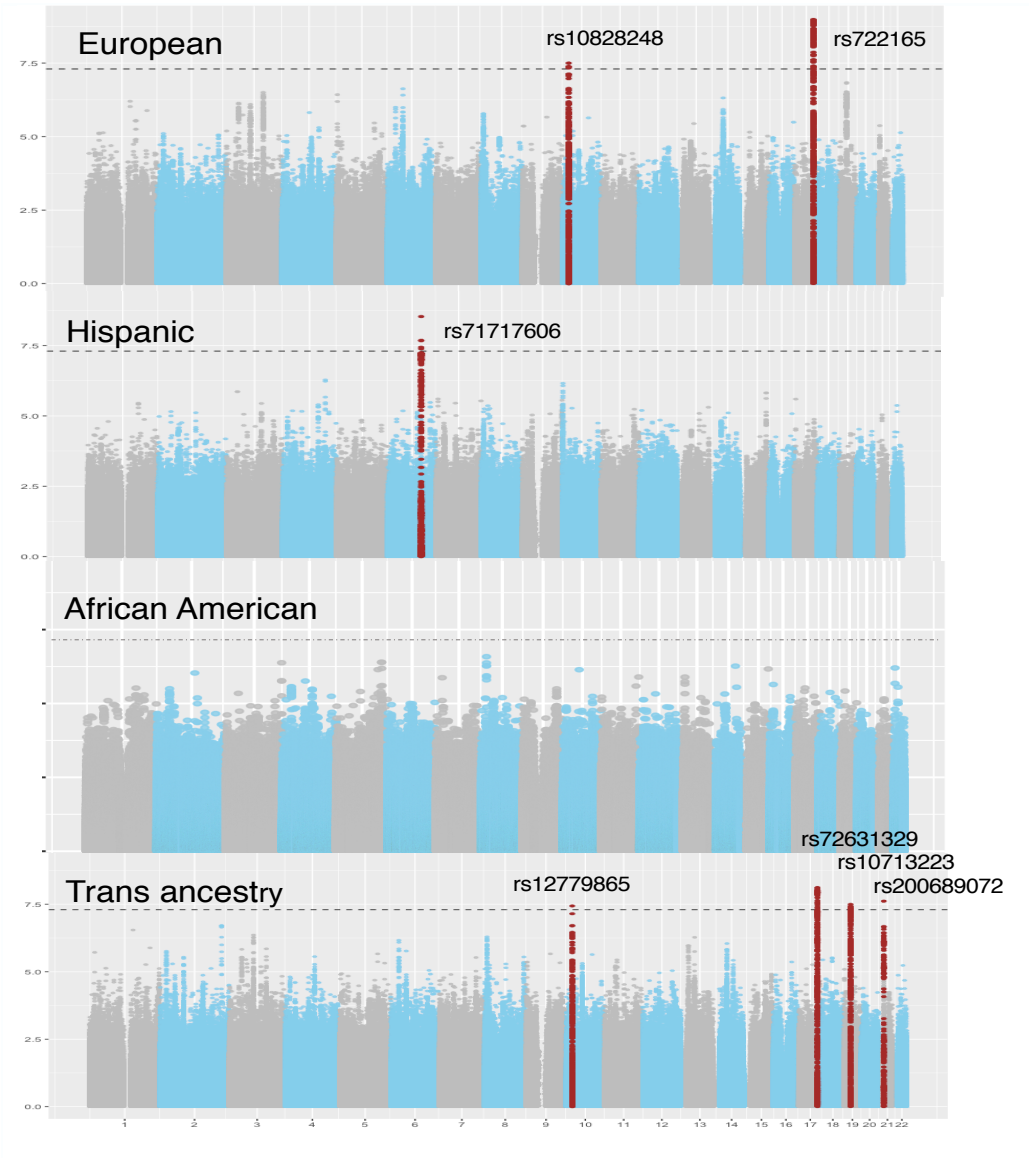
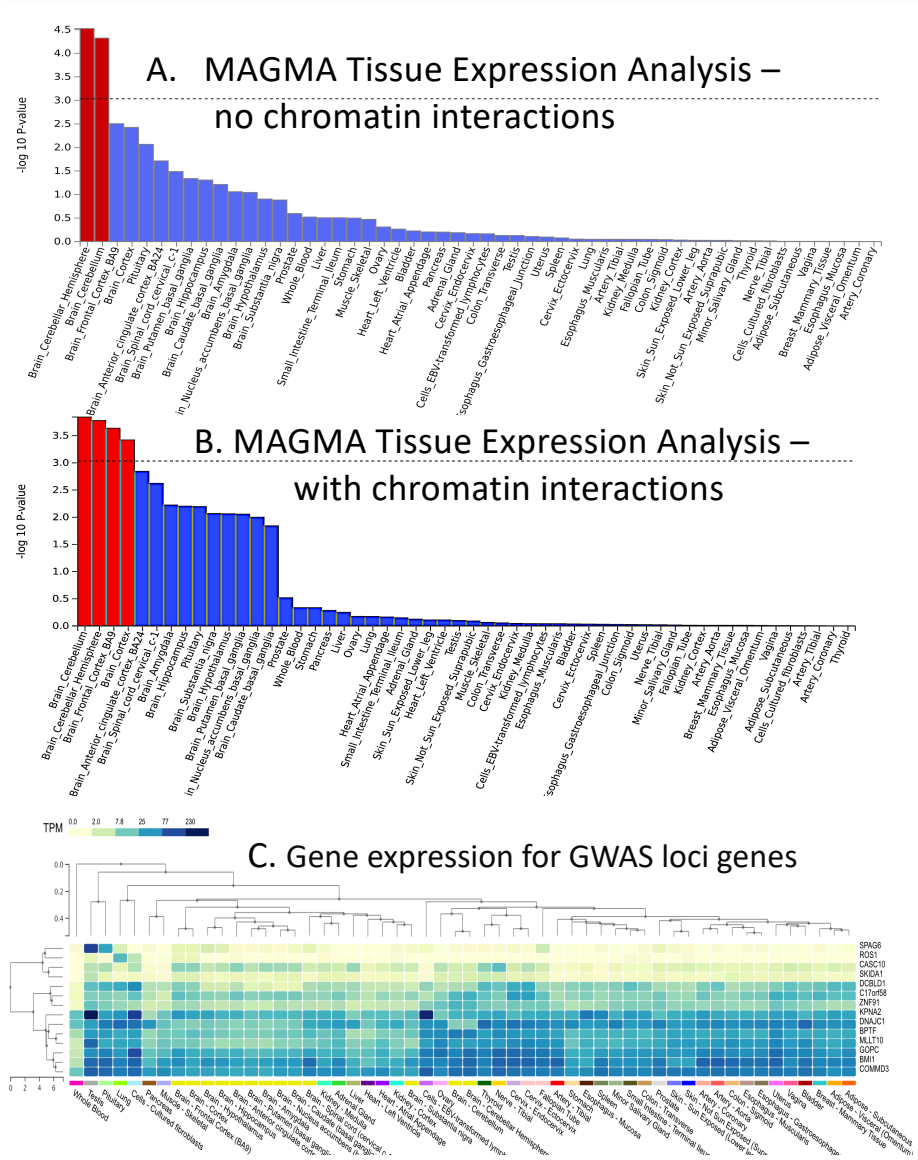


Fig3. Gene enrichment comparing body tissues. A. With no chromatin enhancement, using the full distribution of SNP p-values, only cerebellum shows significant expression. Dotted line denotes significant P-value corrected for 54 tissues. B. With 3-D chromatin analysis including Hi-C and Chia_PET interchromatin analysis, other aspects of the brain are enhanced, specifically frontal cortex. C. Heatmap of gene expression for GWAS loci genes in each GTEx v8 tissue. Rows and columns are arranged by hierarchical clustering of the transcripts per million (TPM) values.



Tables

**Table 1. Demographics**

| | | Cases | (%) | Controls | (%) | Total | (%) | *P*-Value |
|---|---|---|---|---|---|---|---|---|
| TOTAL (%) | | 50,339 | (12.06) | 366,900 | (87.94) | 417,239 | | |
| Age(SD) | | 67 | (11.73) | 61 | (14.62) | 62 | (14.28) | < 0.001 |
| AFR | male | 8,801 | (12.93) | 59,280 | (87.07) | 68,081 | (85.75) | < 0.001 |
| | female | 1,334 | (11.80) | 9,973 | (88.20) | 11,307 | (14.24) | < 0.001 |
| EUR | male | 33,605 | (11.95) | 247,696 | (88.05) | 281,301 | (92.92) | < 0.001 |
| | female | 2,968 | (13.23) | 19,474 | (86.77) | 22,442 | (7.08) | < 0.001 |
| EUR ≥ 50 | | 31,675 | (12.39) | 223,884 | (87.60) | 255,559 | (100.00) | < 0.001 |
| HIS | male | 3,282 | (10.73) | 27,304 | (89.27) | 30,586 | (89.67) | < 0.001 |
| | female | 349 | (9.91) | 3,173 | (90.09) | 3,522 | (10.33) | < 0.001 |

**Table 2. Genome-Wide Significant Loci Associated with Dizziness in the Million Veteran Program with Replication in a European Cohort**

| | | | | Discovery - Million Veteran Program (n = 417,239) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cohort** | **Ch** | **Marker** | **BP[a]** | **A1** | **A2** | **A1Frq** | **INFO** | **BETA** | **SE** | **P-value** | **Z-score** | **Wt.** | **Dir[b]** |
| European | 10 | rs10828248 | 21824619 | A | G | 0.638 | 0.9997 | -0.0458 | 0.008 | 3.17E-08 | 5.532 | 303,743 | |
| | 17 | rs7221651 | 65832390 | A | G | 0.795 | 0.9778 | -0.0623 | 0.010 | 3.71E-10 | 6.266 | 303,743 | |
| Europeans ≥ 50[h] | 10 | rs10828248 | 21824619 | G | A | 0.639 | 0.9997 | -0.0481 | 0.008 | 1.49E-08 | 5.663 | 269,329 | |
| | 17 | rs4791046 | 65851973 | A | G | 0.528 | 0.9778 | -0.0598 | 0.010 | 5.24E-09 | 5.839 | 269,329 | |
| Hispanic | 6 | rs71717606 | 117804875 | C | CT | 0.847 | 0.8871 | -0.2138 | 0.036 | 4.84E-10 | 5.934 | 34,108 | |
| Trans-ancestry[e] | 10 | rs12779865 | 21829162 | T | C | 0.289 | > .60 | 0.0133 | 0.002 | 3.65E-08 | 5.507 | 417,239 | +++ |
| | 17 | rs72631329 | 65818374 | T | C | 0.129 | > .60 | 0.0166 | 0.003 | 7.61E-09 | 5.777 | 417,239 | +++ |
| | 19 | rs10713223 | 23604022 | T | TA | 0.684 | > .60 | -0.0130 | 0.002 | 3.11E-08 | -5.535 | 417,239 | --- |
| | 21 | rs200689072 | 18519809 | A | C | 0.553 | > .60 | -0.0123 | 0.002 | 2.43E-08 | -5.578 | 417,239 | --- |

| | | | | Comparison - (n = 942,613) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Cohort** | **Chr** | **Marker** | **BP[a]** | **A1** | **A2** | **A1Frq** | **OR** | **P-value[c]** | **Dir[d]** | **Predicted Genes within loci** |
| European | 10 | rs10828248 | 21824619 | G | A | 0.3290 | 1.0284 | 4.79E-04 | -- | CASC10; SKIDA1; MLLT10; DNAJC1; COMMD3 COMMD3-BMI1; BMI1; SPAG6 |
| | 17 | rs7221651 | 65832390 | A | G | 0.2000 | 1.0213 | 2.59E-02 | -+ | BPTF; C17orf58; KPNA17 |
| Europeans ≥ 50[h] | 10 | rs10828248 | 21824619 | G | A | 0.3290 | 1.0284 | 4.79E-04 | -+ | CASC10; SKIDA1; MLLT10; DNAJC1; COMMD3; COMMD3-BMI1; BMI1; SPAG6 |
| | 17 | rs4791046 | 65851973 | A | G | 0.2000 | 1.0200 | 3.06E-02 | -+ | BPTF; C17orf58; KPNA17 |
| Hispanic[j] | 6 | rs71717606 | 117804875 | CTTT | CT | 0.0382 | 1.0206 | 2.92E-01 | -+ | GOPC; ROS1; DCBLD1 |
| Trans-ancestry[e] | 10 | rs12779865 | 21829162 | | T | 0.3179 | 1.0237 | 3.98E-03 | ++ | CASC10; SKIDA1; MLLT10; DNAJC1; COMMD3; COMMD3-BMI1; BMI1; SPAG6 |
| | 17 | rs72631329 | 65818374 | | C | 0.1278 | 1.0175 | 1.41E-01 | ++ | BPTF; C17orf58; KPNA17 |
| | 19[f] | rs141771535 | 23601989 | | A | 0.2180 | 1.1031 | 4.55E-27 | [f] | ZNF91; LINC01224 |
| | 21[g] | rs13047989 | 18530070 | | G | 0.3460 | 1.0045 | 3.69E-01 | [g] | |

Abbreviations: A, adenosine; C, cytosine; G, guanidine; T, thymidine; INFO, imputation information score; Frq, frequency; MVP, Million Veteran Program; SE, standard error; OR, odds ratio.

[a] Indicates base pair position on chromosome GrCh37 Human Genome Build hg19.

[b] Comparison of direction of effect in three ancestries in MVP, i.e., European, African American, and Hispanic.

[c] Significance in replication cohort at P-value < 6.25E-03 after Bonferroni correction for 8 variants tested for replication.

[d] Comparison of direction of effect between MVP and replication cohort.

[e] Trans-ancestry consists of European (n = 303,743), African American (n = 79,388) and Hispanic (n = 34,108) meta-analysis

[f] Proxy variant for MVP Marker rs10713223, rs141771535 with *r2* = 0.82.

[g] Proxy variant for MVP Marker rs200689072, rs13047989 with *r2* = 0.66.

[h] MVP participants of European descent ≥ age of 50 (n = 258,533).

[j] Replication in European cohort.

**Table 3. Genome-wide Significant Genes Associated with Dizziness in Million Veteran Program**

| | | | | MVP discovery | | | Iceland/Finland/UKB/US replication | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **GWAS** | **GENE** | **Start[a]** | **Stop[a]** | **# SNPs[b]** | **# params[c]** | **P-value** | **#SNPs[b]** | **# param[c]** | **P-value[d]** | **Dir Dis/Rep** |
| Transancestry | MFHAS1 | 8640864 | 8751155 | 1016 | 101 | 1.14E-06 | 517 | 29 | 0.74379 | +- |
| | BPTF[d] | 65821640 | 65980494 | 832 | 58 | 4.84E-08 | 280 | 13 | 0.03156 | ++ |
| | KPNA2[d] | 66031635 | 66042958 | 78 | 20 | 1.03E-06 | 39 | 5 | 0.33328 | ++ |
| | ZNF91[d] | 23487793 | 23578362 | 584 | 51 | 4.98E-07 | *304 | 12 | 6.14E-14 | ++ |
| European | MFHAS1 | 8640864 | 8751155 | 550 | 28 | 1.02E-06 | 517 | 29 | 0.74379 | ++ |
| | MLLT10[d] | 21823094 | 22032559 | 193 | 23 | 1.78E-06 | *208 | 32 | 0.00176 | **++** |
| | BPTF[d] | 65821640 | 65980494 | 305 | 9 | 5.04E-10 | 280 | 13 | 0.03156 | ++ |
| | C17orf58[d] | 65987217 | 65989765 | 3 | 1 | 2.54E-07 | 3 | 1 | 0.02915 | ++ |
| | KPNA2[d] | 66031635 | 66042958 | 39 | 4 | 6.76E-07 | 39 | 5 | 0.33328 | ++ |
| | ZNF91[d] | 23487793 | 23578362 | 329 | 11 | 1.58E-06 | *304 | 12 | 6.14E-14 | ++ |
| Eur >= 50 | SKIDA1[d] | 21802407 | 21814611 | 11 | 4 | 1.92E-06 | 11 | 5 | 0.00994 | ++ |
| | MLLT10[d] | 21823094 | 22032559 | 193 | 23 | 4.97E-07 | *208 | 32 | 0.00176 | **++** |
| | BPTF[d] | 65821640 | 65980494 | 306 | 9 | 6.51E-09 | 280 | 13 | 0.03156 | ++ |
| Hispanic | GOPC | 117639374 | 117923691 | 1157 | 46 | 2.44E-07 | 1053 | 47 | 0.81116 | ++ |

Abbreviations: MVP, Million Veteran Program; CHR chromosome; Dir Dis/Rep, direction of effect, discovery/replication

[a] Start position in GrCh37 Human Genome Build 19.

[b] SNPs tested within gene.

[c] Number of variants used to compute the gene-wise statistics.

[d] Significance in replication cohort at P < 6.25E-03 for 8 genes tested.

**Table 4. Fine Mapping Credible Sets**

| SNP | CHR | BP[1] | $r_g^2$ | ALLELE1 | ALLELE0 | A1FREQ | P | BETA | SE |
|---|---|---|---|---|---|---|---|---|---|
| rs72965321 | 6 | 117388218 | [3] | C | T | 0.944778 | 1.71E-05 | -0.22678 | 0.051611 |
| rs2883416 | 6 | 117676330 | [3] | C | T | 0.060377 | 4.40E-03 | 0.165584 | 0.057228 |
| rs7894565 | 10 | 21583984 | 0.7428 | T | C | 0.678874 | 2.56E-07 | -0.04545 | 0.008803 |
| rs11322024 | 10 | 21747989 | 0.7609 | TC | T | 0.379806 | 4.63E-06 | 0.045273 | 0.009866 |
| rs9899520 | 17 | 67961452 | 1.0 | A | G | 0.228443 | 2.10E-09 | 0.059577 | 0.009907 |
| rs7221651 | 17 | 67836274 | 1.0 | G | A | 0.795488 | 3.71E-10 | -0.06228 | 0.009897 |
| rs1976054 | 17 | 67837108 | 1.0 | T | G | 0.795494 | 3.73E-10 | -0.06227 | 0.009897 |
| rs7208909 | 17 | 67837430 | 1.0 | T | C | 0.795494 | 3.73E-10 | -0.06227 | 0.009897 |
| rs6504542 | 17 | 67840526 | 1.0 | T | C | 0.7955 | 3.72E-10 | -0.06228 | 0.009897 |
| rs612285 | 19 | 23421337 | 0.9075 | A | G | 0.7009 | 3.78E-08 | *-5.501 | *-5.501 |
| rs620916 | 19 | 23414408 | 0.8971 | T | C | 0.2992 | 4.21E-08 | *5.482 | *5.482 |

Abbreviations: SNP, single nucleotide polymorphism; CHR, chromosome, POS, position; $r_g$, $R^2$ correlation with lead SNP in locus; ALLELE0, reference allele; ALLELE1, effect allele; SE, standard error.

[1] Analyzed in ancestry-specific 1000 genome sets, GrCh38.

[2] Linkage disequilibrium score with lead SNP in the locus. All P-values for correlations < 0.001. Data from LDLink, https://ldlink.nci.nih.gov/ .

[3] Lead SNP not in reference sets for 1000 genomes.