The American Journal of Human Genetics, Volume *108*

Supplemental information

Phenome risk classification enables phenotypic imputation and gene discovery in developmental stuttering

Douglas M. Shaw, Hannah P. Polikowsky, Dillon G. Pruett, Hung-Hsin Chen, Lauren E. Petty, Kathryn Z. Viljoen, Janet M. Beilby, Robin M. Jones, Shelly Jo Kraft, and Jennifer E. Below

Supplemental Figures

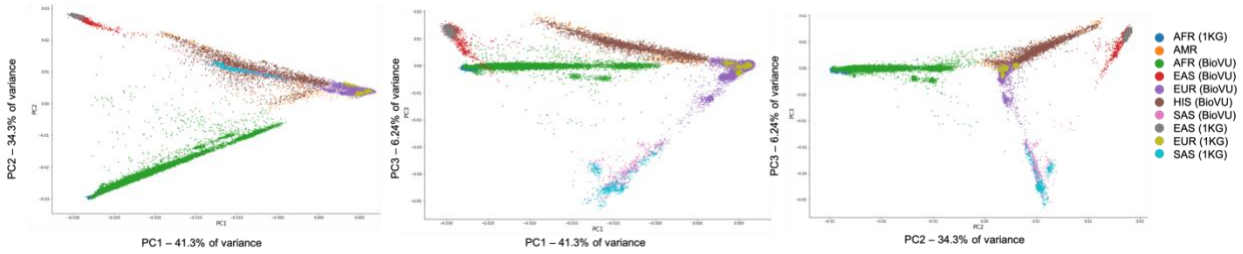

Figure S1. Principal component analysis of BioVU patients results. Top three principal components for all subjects in BioVU projected onto 1KG reference data. Broad ancestry groups were stratified into either African (AFR), European (EUR), Hispanic (HIS), South East Asian (SAS), East Asian (EAS) ancestry, or admixed Americans (AMR) based on PCs 1-3 (see methods). EUR ancestry was stratified

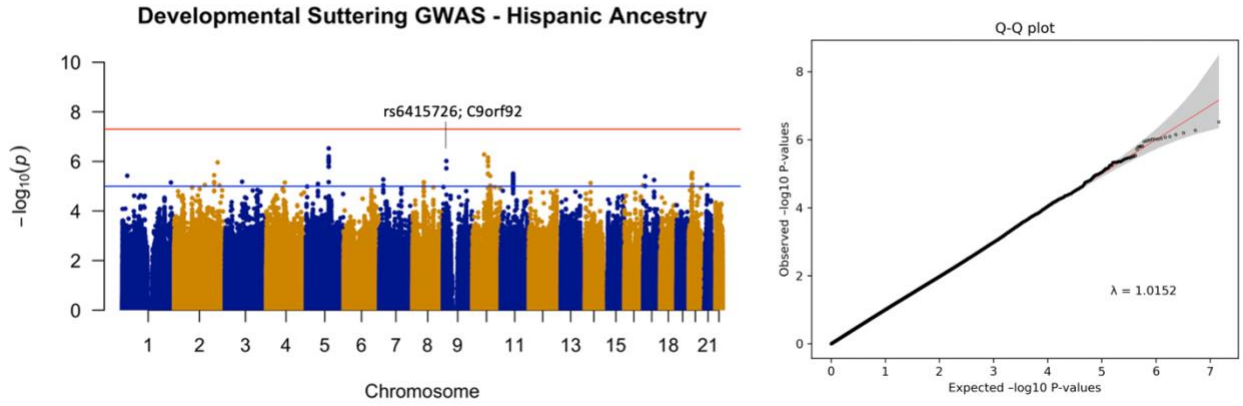

Figure S2. Manhattan and qq-plot of Hispanic ancestry PheML predicted developmental stuttering GWAS results. Analysis included 8,147,169 autosomal variants. No variants reached genome-wide significance ($P<5*10^{-8}$). Red line indicates genome-wide significance threshold ($5.0*10^{-8}$), blue line indicates suggestive significance threshold ($1.0*10^{-5}$). Loci reported on table 2 are labeled on plot.

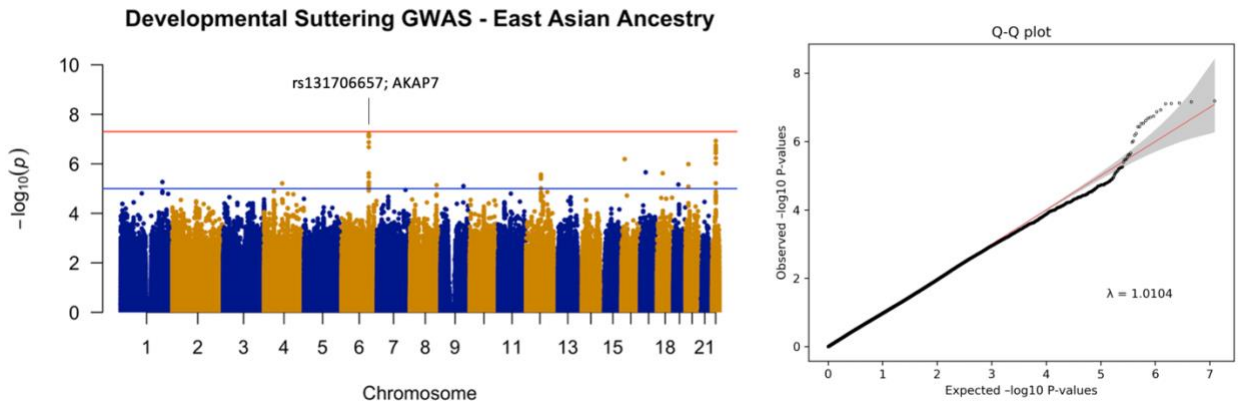

Figure S3. Manhattan and qq-plot of East Asian ancestry PheML predicted developmental stuttering GWAS results. Analysis included 6,922,517 autosomal variants. No variants reached genome-wide significance ($P<5*10^{-8}$). Red line indicates genome-wide significance threshold ($5.0*10^{-8}$), blue line indicates suggestive significance threshold ($1.0*10^{-5}$). Loci reported on table 2 are labeled on plot.

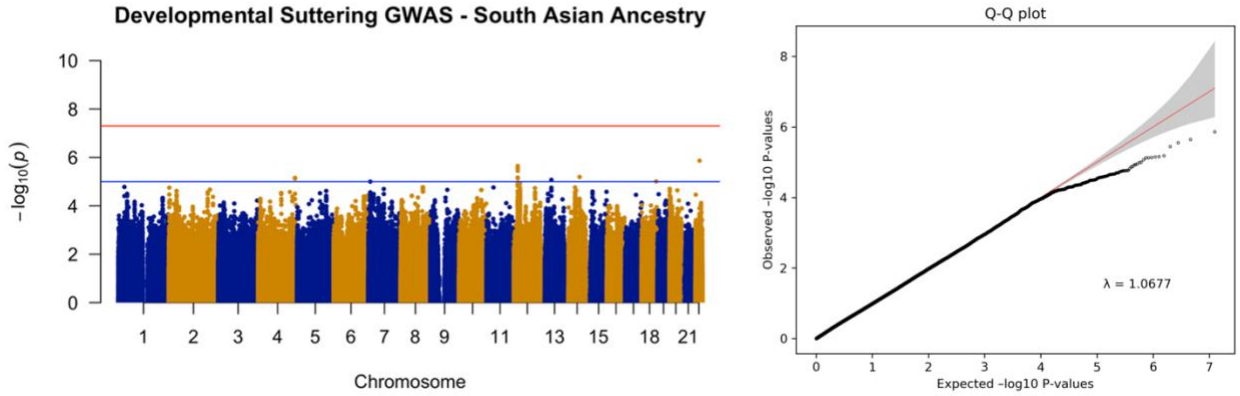

Figure S4. Manhattan and qq-plot of South Asian PheML predicted developmental stuttering GWAS results. Analysis included 7,058,354 autosomal variants. No variants reached genome-wide significance ($P<5*10^{-8}$). Red line indicates genome-wide significance threshold ($5.0*10^{-8}$), blue line indicates suggestive significance threshold ($1.0*10^{-5}$). Only 51 subjects of SAS ancestry were predicted by the PheML model to have developmental stuttering.

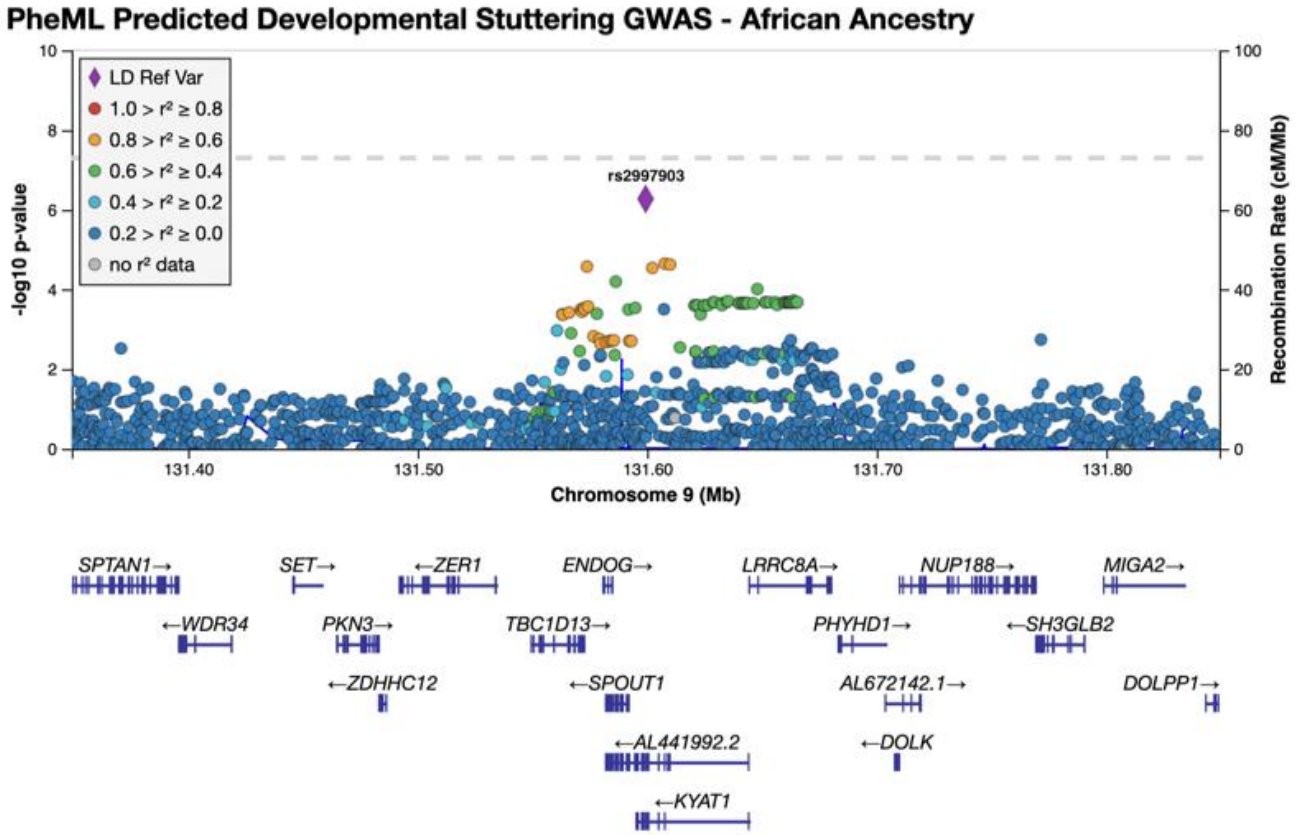

Figure S5. LocusZoom Plot for rs2997903 in AFR PheML Stuttering GWAS. Lead variant found within the first intron of *KYAT1* (beta=0.308; *P*=5.32*10^{-7}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

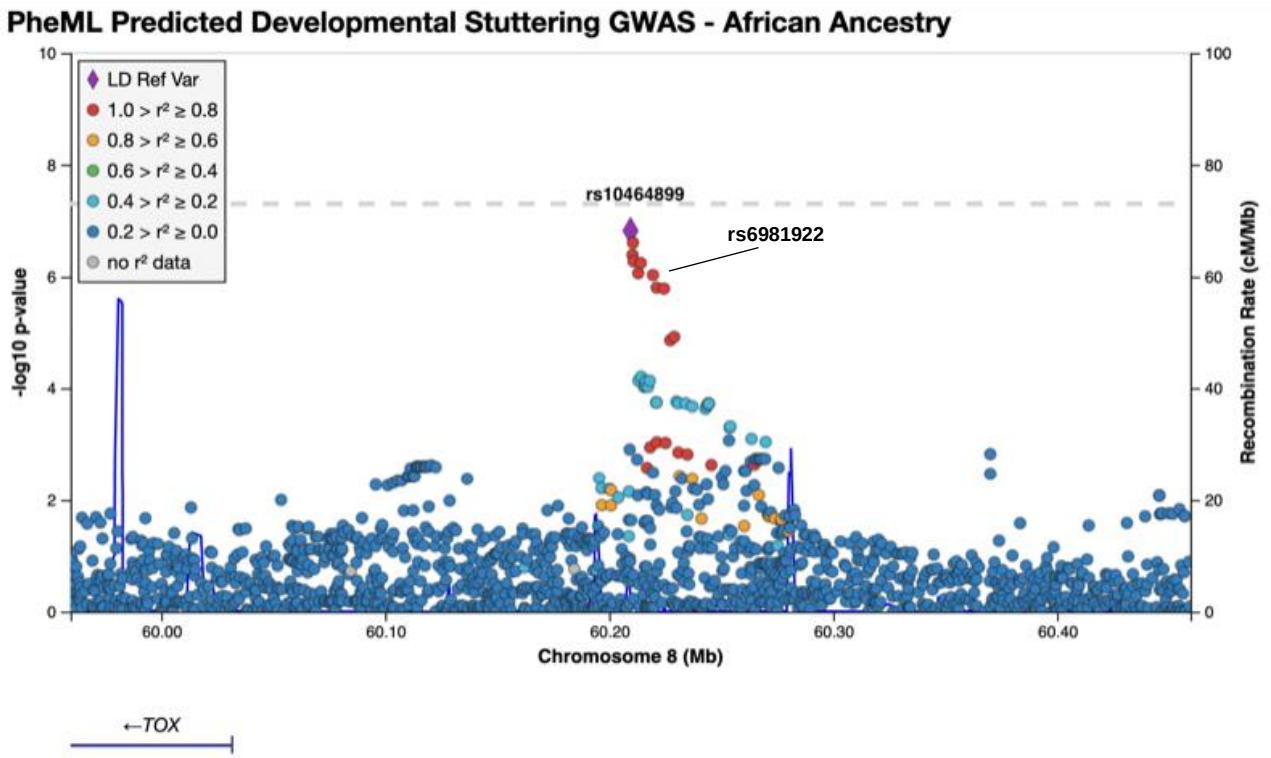

Figure S6. LocusZoom Plot for rs10464899 in AFR PheML Stuttering GWAS. Lead variant found 178kb 5' of *TOX* (beta=0.216; P=1.51*10^{-7}). We also reported rs6981922 (beta=0.197; P=9.35*10^{-7}) which replicated in our East Asian population as well (P=3.27*10^{-2}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

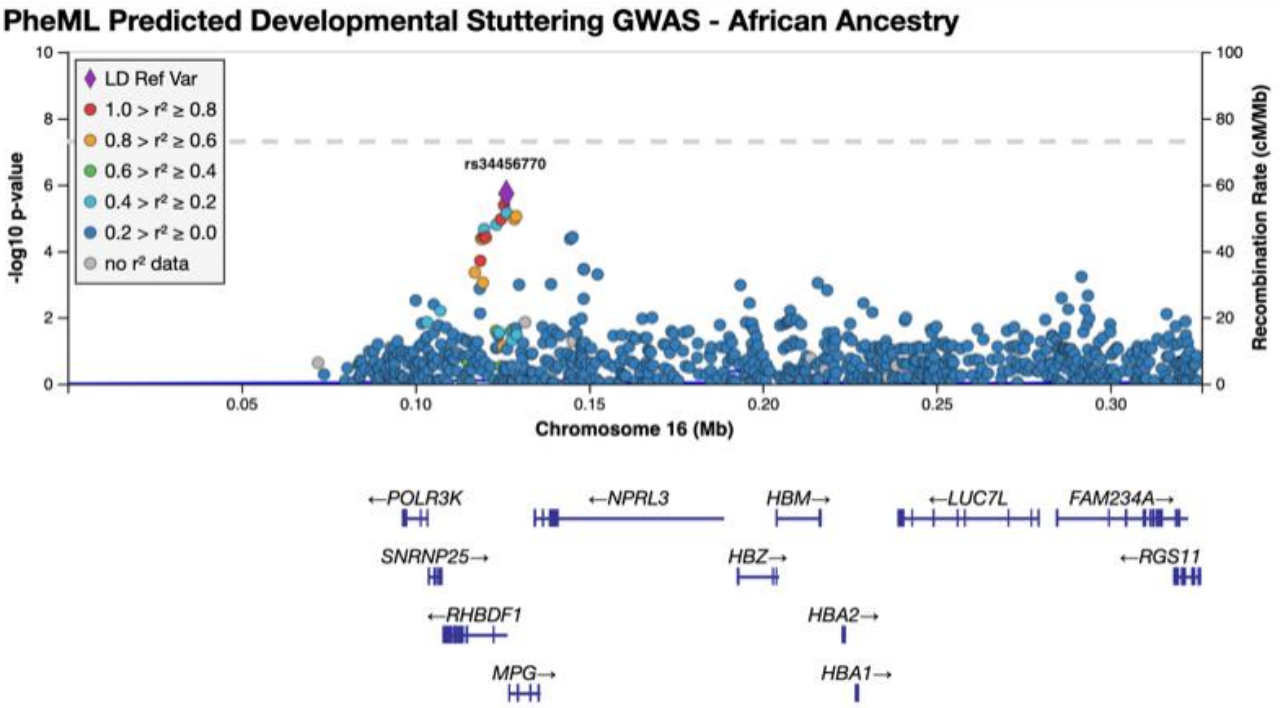

Figure S7. LocusZoom Plot for rs34456770 in AFR PheML Stuttering GWAS. Lead variant found 797bp 5' of *MPG* (beta=0.256; P=1.87*10^{-6}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

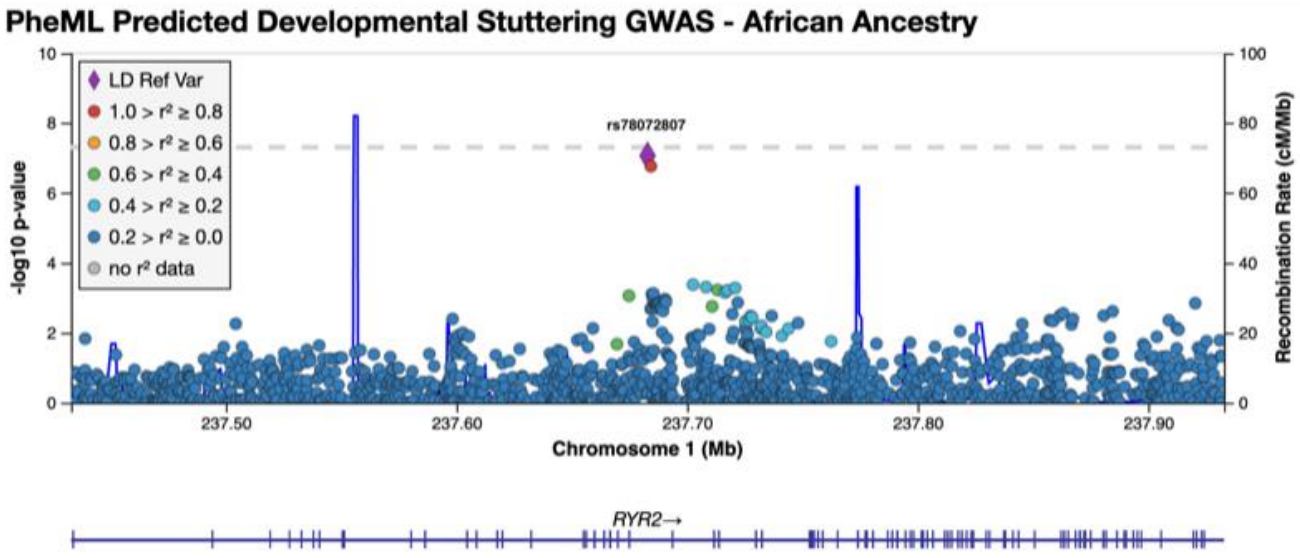

Figure S8. LocusZoom Plot for rs78072807 in AFR PheML Stuttering GWAS. Lead variant found within the 21st intron of *RYR2* (beta=0.371; P=8.73*10^{-8}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

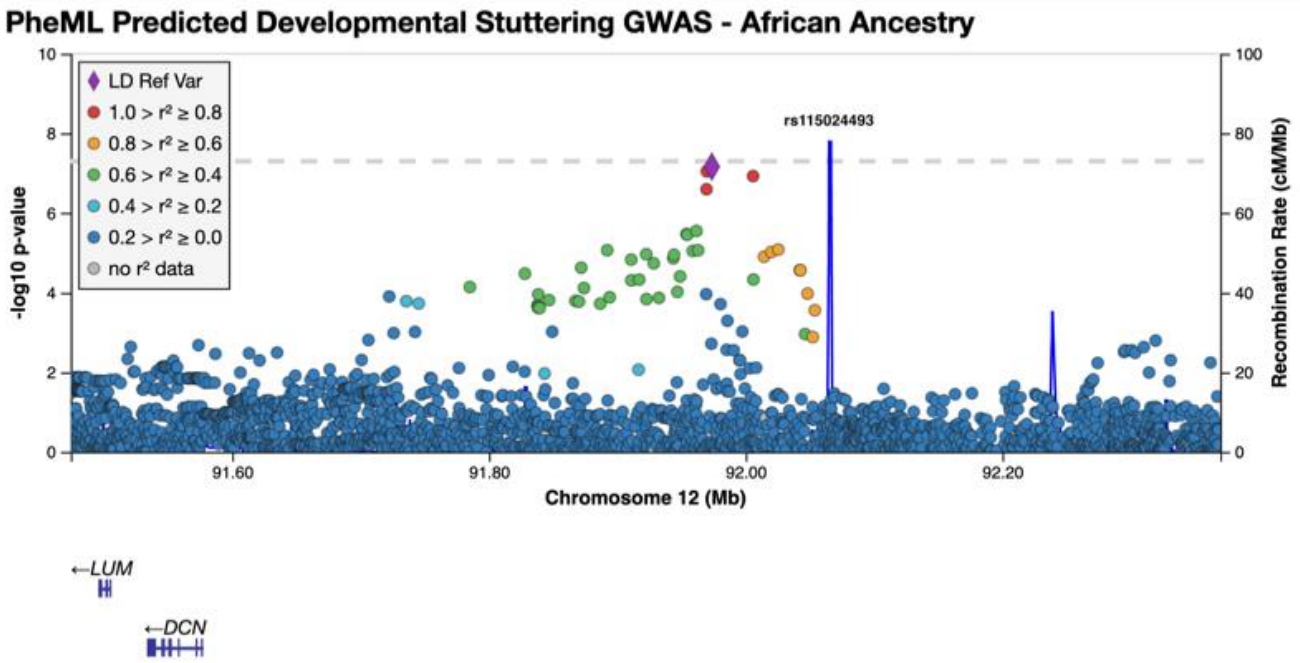

Figure S9. LocusZoom Plot for rs115024493 in AFR PheML Stuttering GWAS. Lead variant found 397kb 5' of *DCN* (beta=0.376; *P*=6.58*10^{-8}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

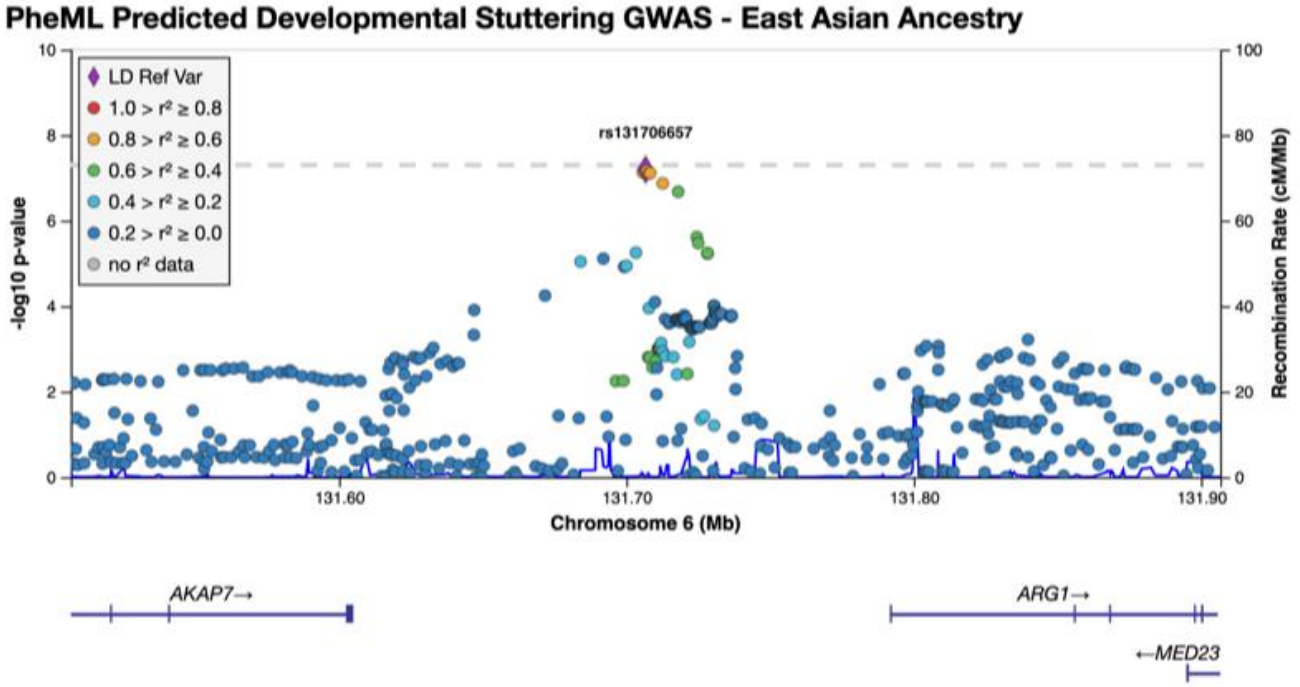

Figure S10. LocusZoom Plot for rs10872381 in EAS PheML Stuttering GWAS. Lead variant found 102kb 3' of *AKAP7* (beta=0.803; *P*=6.38*10^{-8}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

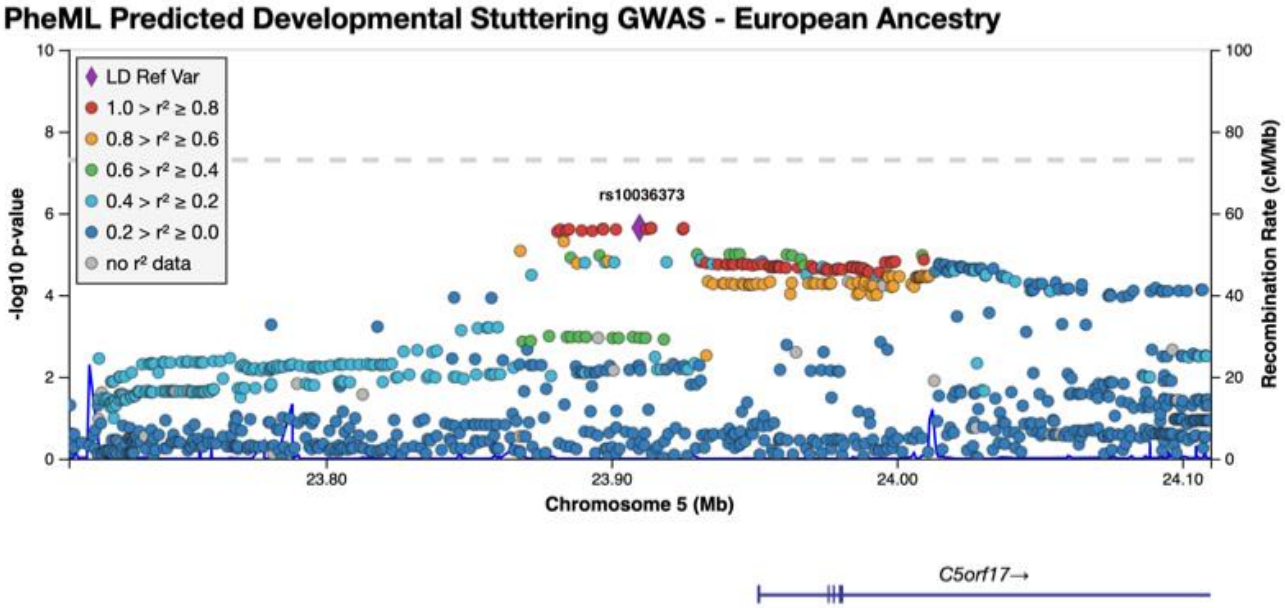

Figure S11. LocusZoom Plot for rs10036373 in EUR PheML Stuttering GWAS. Lead variant found 42kb 5' of *C5orf17* (beta=0.701; *P*=3.68*10^{-6}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

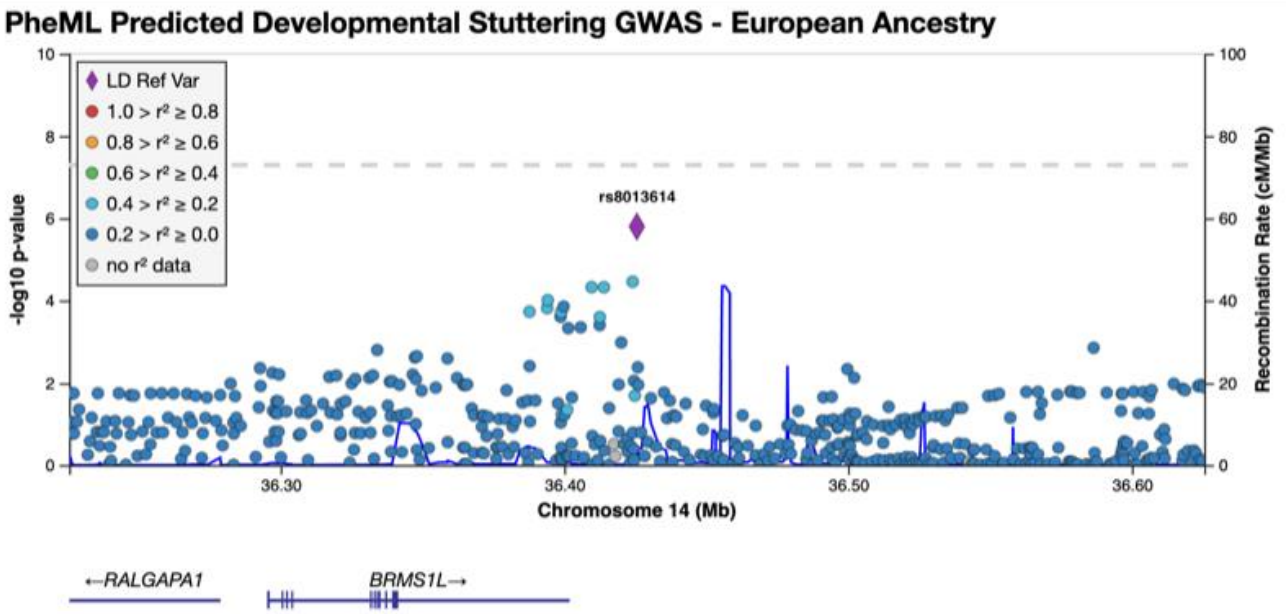

Figure S12. LocusZoom Plot for rs8013614 in EUR PheML Stuttering GWAS. Lead variant found 84kb 5' of *BRMS1L* (beta=-.120; *P*=1.59*10^{-6}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

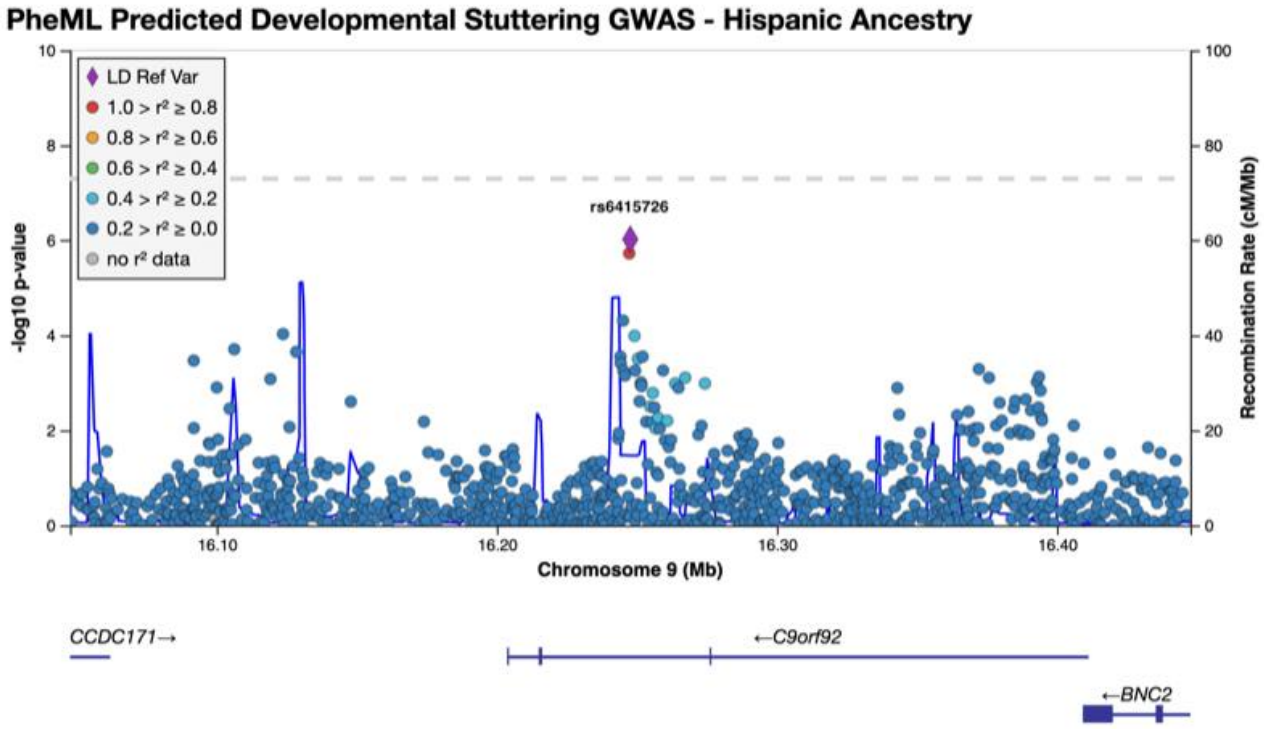

Figure S13. LocusZoom Plot for rs6415726 in HIS PheML Stuttering GWAS. Lead variant found within the second intron of *C9orf92* (beta=0.730; *P*=9.61*10^{-6}). Dashed line indicates genome-wide significance threshold (5.0*10^{-8}).

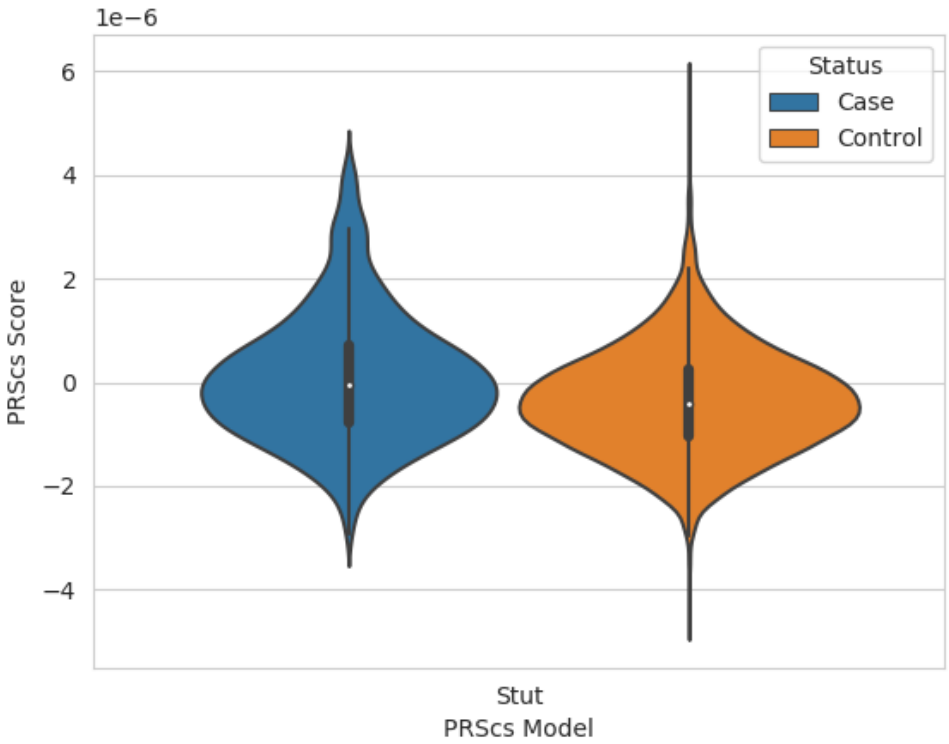

Figure S14. Polygenic risk score violin plots. PRS model was developed using the summary statistics from the EUR ancestry PheML stuttering GWAS. The International Stuttering Project stuttering case set (blue) scored significantly higher on the PRS model (mean=$8.56*10^{-8}$, SD=$1.13*10^{-6}$) than their matched controls (orange), (mean=$-3.59*10^{-7}$, SD = $1.01*10^{-6}$; *t(1131)*=13.12, *P* = $6.83*10^{-39}$).

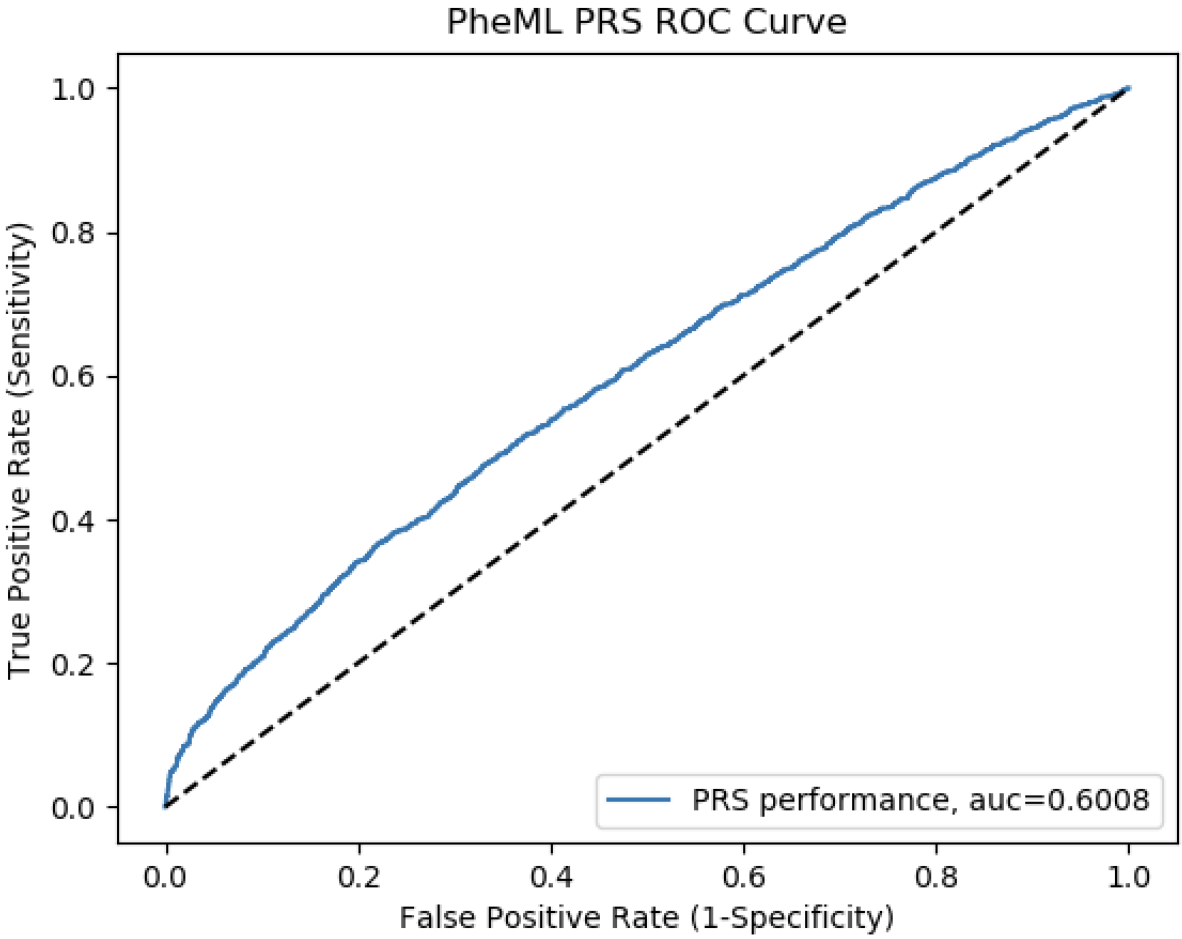

Figure S15. Polygenic risk score receiver operating characteristic (ROC) curve. PRS model was developed using the summary statistics from the EUR ancestry PheML stuttering GWAS. ROC curve plotted to demonstrate the model performance in predicting stuttering liability in the International Stuttering Project stuttering set. Area under the curve (AUC) = 0.60.

Supplemental Tables

Supplementary table S1. ICD codes used to identify developmental stuttering patients		
ICD-9 Code	ICD-10 Code	Definition
307.0	F98.5	Adult-Onset Fluency Disorder
315.35	F80.81	Childhood Onset Fluency Disorder
784.5	R47.82	Fluency Disorder in Conditions Classified Elsewhere

Table S1. ICD codes used to identify developmental stuttering.

Supplementary table S2. Demographics of clinically validated stuttering case and control set

	Stuttering Cases		Population Controls
Total	1345		7019
Male	965 (71.7%)		4951 (70.5%)
Female	380 (28.3%)		2068 (29.5%)
Ancestry		n (%)	
European	1132 (84.2%)		6111 (87.1%)
African	68 (5.1%)		400 (5.7%)
East Asian	42 (3.1%)		116 (1.7%)
South Asian	44 (3.3%)		148 (2.1%)
Hispanic	38 (2.8%)		132 (1.9%)
Mixed/Other	21 (1.5%)		112 (1.6%)

Table S2. Demographic distribution for subjects used in genome-wide association analysis for the International Stuttering Project (ISP) stuttering sample set.

See attached table

Table S3. Suggestive hits from PheML predicted developmental stuttering GWAS run in each ancestry. Table includes all variants where $P<5.0*10^{-6}$. We also report association results for each variant in alternative ancestries. GWAS results for European, African, Hispanic, South Asian, and East Asian ancestry cohorts denoted as EUR, AFR, HIS, SAS, and EAS respectively. GWAS results from clinically validated set denoted as CV.

See attached table

Table S4. Replication results of previously identified genes associated with stuttering.