HGGA, Volume *3*

Supplemental information

Population-based genetic effects

for developmental stuttering

Hannah G. Polikowsky, Douglas M. Shaw, Lauren E. Petty, Hung-Hsin Chen, Dillon G. Pruett, Jonathon P. Linklater, Kathryn Z. Viljoen, Janet M. Beilby, Heather M. Highland, Brandt Levitt, Christy L. Avery, Kathleen Mullan Harris, Robin M. Jones, Jennifer E. Below, and Shelly Jo Kraft

Supp. Table 1. ICD codes used to identify developmental stuttering patients in BioVU		
ICD-9 Code	ICD-10 Code	Definition
315.35	F80.81	Childhood Onset Fluency Disorder
315.39	F80.0, F80.89, F80.9	Other developmental speech or language disorder

Table S1. ICD9/10 Exclusion criteria. Table depicts ICD codes applied as exclusion criteria for the selection of ancestry- and sex-matched population-based controls for the ISP GWAS of clinically ascertained cases.

rsID	CHR	Position	Effect Allele	Other Allele	EAF	INFO	BETA	SE	OR	P VALUE	Nearest gene	Location
NA	2	78519487	T	TA	0.343	0.991	-0.270	0.051	0.763	1.08E-07	NA	NA
NA	4	172506840	C	CTACTC ATGAG	0.067	0.682	0.587	0.116	1.80	4.23E-07	*GALNTL6*	intronic
rs61859204	10	43660815	T	C	0.095	0.989	0.407	0.081	1.50	5.02E-07	NA	NA
rs7197535	16	24382574	G	A	0.1144	0.982	0.370	0.074	1.45	6.20E-07	*CACNG3*	20kb downstream
rs10994295	10	46040909	C	T	0.499	0.487	0.418	0.084	1.512	6.42E-07	*MSMB*	intronic
rs1045247	1	226994874	A	G	0.180	0.985	0.302	0.062	1.352	1.05E-06	*CDC42BPA*	exonic
rs6427338	1	156953690	C	A	0.037	0.705	0.730	0.150	2.08	1.13E-06	*ARHGEF11*	intronic
rs904872	15	97520985	T	C	0.071	0.402	0.715	0.148	2.04	1.35E-06	NA	NA
rs144946999	2	25505099	G	A	0.012	0.972	1.07	0.223	2.90	1.71E-06	*DTNB*	intronic
rs10817935	9	116790777	T	C	0.234	0.971	-0.278	0.059	0.758	2.35E-06	*ASTN2*	intronic
rs2774497	13	103286349	C	T	0.086	0.971	-0.413	0.089	0.661	3.19E-06	*SLC10A2*	220kb downstream
rs182338960	14	85216697	G	A	0.021	0.916	0.817	0.176	2.265	3.24E-06	*SNORD3P3*	55kb upstream
rs9531835	13	31400237	C	G	0.188	0.890	0.301	0.065	1.351	3.42E-06	*B3GLCT*	68kb downstream
rs34711070	4	182770462	C	G	0.482	0.888	0.239	0.052	1.27	3.61E-06	*TENM3*	intronic
rs74434153	4	33128906	C	T	0.045	0.882	0.556	0.121	1.74	4.01E-06	NA	NA
rs13183297	5	1200770	G	C	0.232	0.917	0.275	0.060	1.32	4.19E-06	*SLC6A19*	1kb upstream
rs12613469	2	131920387	A	C	0.146	0.936	-0.342	0.074	0.710	4.30E-06	*CDRT15P3*	119kb downstream
rs549259468	4	118420869	G	C	0.012	0.970	1.01	0.221	2.75	4.90E-06	*PRSS12*	68kb downstream
rs71571436	6	166126728	T	C	0.030	0.985	-0.703	0.154	0.495	4.91E-06	*TBXT*	31kb upstream

Table S3. Top hits from ISP genome wide association study. Table includes nineteen sentinel variants from loci with $p < 5 \times 10^{-6}$. NA (not available) reported for variants where the nearest protein-coding gene was more than 250kb away (either upstream or downstream according to UCSC reference genome browser). Location indicates approximate position from nearest gene. Base-pair positions listed according to human genome reference build 38.

rsID	CHR	Position	Effect Allele	Other Allele	EAF	INFO	BETA	SE	OR	P VALUE	Nearest gene	Location
rs76016608	21	36661116	T	G	0.082	0.937	0.530	0.102	1.70	2.08E-07	*RUNX1*	240kb downstream
rs77605856	5	63572220	G	T	0.062	0.508	0.799	0.154	2.22	2.18E-07	*RNF180*	intronic
rs113284510	3	8725187	T	C	0.079	0.478	-0.787	0.152	0.455	2.23E-07	*SSUH2*	intronic
NA	12	29759906	G	GCC	0.380	0.970	0.288	0.056	1.33	2.27E-07	*TMTC1*	intronic
rs61936749	12	95114780	T	C	0.081	0.876	0.551	0.107	1.74	2.35E-07	*TMCC3*	70kb downstream
rs57878561	2	52386461	C	G	0.087	0.724	0.565	0.114	1.76	6.79E-07	NA	NA
rs3098526	15	27840730	C	T	0.461	0.862	-0.289	0.058	0.749	7.27E-07	*GABRG3*	62kb downstream
rs11193269	10	108998421	T	A	0.080	0.872	0.523	0.106	1.69	7.69E-07	*SORCS1*	74kb downstream
rs58528263	4	17107477	G	T	0.226	0.991	0.329	0.067	1.39	9.13E-07	*LDB2*	207kb downstream
rs1491167301	1	173738919	C	CA	0.257	0.456	0.447	0.091	1.56	9.19E-07	*KLHL20*	intronic
rs915383	14	101651628	T	C	0.067	0.563	0.687	0.142	1.99	1.23E-06	NA	NA
rs73479865	9	25555477	C	T	0.026	0.864	0.818	0.173	2.27	2.12E-06	*TUSC1*	121kb upstream
rs73859317	3	106215273	G	C	0.031	0.896	0.770	0.163	2.16	2.29E-06	NA	NA
rs78908171	8	70551392	A	G	0.114	0.806	0.452	0.096	1.57	2.49E-06	*SULF1*	intronic
rs6824924	4	38761762	G	T	0.109	0.543	-0.556	0.119	0.573	2.92E-06	*TLR10*	12kb upstream
rs35164368	9	100375608	G	A	0.046	0.875	0.648	0.139	1.92	3.17E-06	*TSTD2*	intronic
rs2070394	21	35239405	T	A	0.324	0.979	0.272	0.059	1.31	3.69E-06	*ITSN1*	intronic
rs151193015	1	173441135	C	G	0.034	0.647	0.865	0.188	2.38	4.07E-06	NA	NA
rs28380337	5	153785436	T	G	0.064	0.910	0.560	0.122	1.75	4.31E-06	NA	NA
rs564196491	11	58977318	G	del(TATGTGTGTGCGTGTGTGTGTGTG)	0.388	0.640	-0.312	0.068	0.732	4.35E-06	*MPEG1*	3' UTR
rs3786426	18	9792798	T	C	0.144	0.544	0.466	0.101	1.59	4.40E-06	*RAB31*	intronic
NA	7	125074119	A	AT	0.098	0.746	0.480	0.105	1.62	4.49E-06	NA	NA
rs2868820	4	84811756	A	G	0.478	0.750	0.282	0.062	1.33	4.86E-06	NA	NA
rs2433647	12	23018273	G	A	0.341	0.980	0.265	0.058	1.30	4.90E-06	*ETNK1*	175kb downstream

Table S4. Top hits from Add Health stuttering genome wide association study. Table includes twenty-four sentinel variants from loci with $p < 5 \times 10^{-6}$. NA (not available) reported for variants where the nearest protein-coding gene was more than 250kb away (either upstream or downstream according to UCSC reference genome browser). Location indicates approximate position from nearest gene. Base-pair positions listed according to human genome reference build 37.

Study	Gene	Chr:bp (Gene)	Population Architecture	N SNPs	Lowest pval in our study	effN tests	Adj pval	Rsid	Chr:pos (Rsid)	Effect Allele	Other Allele	EAF	Zscore
Lan, et al. 2009[31]	*DRD2*	chr11:113,409,605-113,475,398	Case-control (gender-matched controls) association study focusing specifically on dopaminergic gene haplotypes and allele frequencies among SNPs in the Han Chinese population	167	0.034	22	0.748	rs12805897	11:113453571	A	G	0.063	2.12
Kang, et al. 2010[32]	*GNPTAB*	chr12:101,745,499-101,830,959	77 unrelated Pakistani individuals that stutter plus unrelated cases from 46 Pakistani families (see Suresh et al.), 270 affected unrelated individuals from N. America & England	199	0.013	14	0.182	rs55764824	12:101788264	T	C	0.087	-2.48
	GNPTG	chr16:1,351,931-1,364,113		19	0.015	7	0.105	rs761057	16:1354725	T	C	0.099	2.43
	NAGPA	chr16:5,024,844-5,034,141		28	0.022	6	0.132	rs34742205	16:5030890	A	G	0.105	-2.29
Raza, et al. 2015[33]	*AP4E1*	chr15:50,908,683-51,005,895	Camroonian family + probands from Cameroon, Pakistani, and N. America	224	0.126	9	1	rs79933276	15:50928230	T	A	0.024	-1.53
Mohammadi, et al. 2017[30]	*CYP17A1*	chr10:102,830,531-102,837,472	Case-control (age and ethnic background matched) study of Kurdish population aged 3-9 years from western Iran	17	0.025	4	0.1	rs4919687	10:102835491	A	G	0.27	-2.24

Table S5. Summary of genes implicated in developmental stuttering from the literature. Reported p value represents the lowest p value identified in the gene found within our meta-GWAS. Adjusted p value (Adj pval) represents the Bonferroni corrected p value after adjusting for the effective number of tests (effN tests). Chromosome and base-pair positions listed according to human genome reference build 38.

Region	Gene	RCP	# of SNPs	Lead SNP	Max SCP	Tissue
chr2:111630529-112630529	*FBLN7*	0.092	10	**rs10779884**	0.030	Muscle Skeletal
chr2:111630529-112630529	*FBLN7*	0.063	13	rs4849044	0.025	Muscle Skeletal
chr2:111630529-112630529	*FBLN7*	0.062	16	**rs10779884**	0.025	Esophagus Mucosa
chr2:111630529-112630529	*FBLN7*	0.060	14	rs6707397	0.023	Brain Hypothalamus
chr2:60940832-61940832	NA	0.052	7	rs74181273	0.046	Esophagus Muscularis
chr2:60940832-61940832	NA	0.051	5	rs74181273	0.033	Esophagus Gastroesophageal Junction
chr2:97986856-98986856	*INPP4A*	0.050	4	rs140321250	0.018	Esophagus Mucosa

Table S6. Identified genomic regions that showed the highest probability of a colocalized association signal between *cis*-eQTLs in GTEx_v8 and stuttering meta-analysis hits. A region is listed if its regional colocalization probability (RCP) is ≥ 0.05. Max SCP (SNP-level colocalization probability) indicates the maximum probability reported for each SNP identified in the indicated region. The affected protein-coding gene is listed; "NA"

indicates no identified protein-coding gene. The tested tissue indicated in the tissue column. SNP in bold was identified as a top hit in the meta-analysis. Region positions refer to human genome reference build 38.

Genes	Gene Count	Adj pval	pval	Tissue	Biological Process
SSUH2*/LRRTM1/****FBLN7****/ELOVL2/RELN/MAB21L2/SEMA6D*	7	6.63E-03	5.10E-04	Minor Salivary Gland	Extracellular matrix and structure organization, cell adhesion, anatomical structure development, developmental process, nervous system development, ossification, neurogenesis, cell migration, bone morphogenesis
PSAT1*/SMURF1/PDE6D/ELOVL2/ ERBB4/RASSF8/NFATC2/ZNF496/ DCTN4/C4orf19/RLBP1*	11	0.123	0.012	Brain Cortex	organic acid catabolic process, carboxylic acid catabolic process, small molecule catabolic process, monocarboxylic acid catabolic process, fatty acid catabolic process, organic acid metabolic process, small molecule metabolic process, oxoacid metabolic process
QDPR/PLLP/TMCC3	3	0.137	0.019	Cerebellar Hemisphere	ensheathment of neurons, axon ensheathment, myelination, neurogenesis, oligodendrocyte differentiation, glial cell differentiation, nervous system development, generation of neurons
USP34*/RELN/C14orf132/EXOSC6 /FRY*	5	0.137	0.025	Cerebellar Hemisphere	DNA binding, transcription regulator activity, metal ion binding, nucleic acid binding
DSEL/QDPR/PLLP/TMCC3 /EFNA1	5	0.167	0.033	Brain Cortex	ensheathment of neurons, axon ensheathment, myelination, neurogenesis, oligodendrocyte differentiation, glial cell differentiation, nervous system development, generation of neurons
QDPR/SLC48A1/PLLP/TMCC3	4	0.178	0.023	Cerebellum	ensheathment of neurons, axon ensheathment, myelination, neurogenesis, oligodendrocyte differentiation, glial cell differentiation, nervous system development, generation of neurons
INPP4A/USP34*/TBC1D30/RELN /UNC13C/C14orf132/FRY/SACM1L*	8	0.178	0.040	Cerebellum	DNA binding, transcription regulator activity, metal ion binding, nucleic acid binding
DSEL/QDPR/CDC42BPA/SEMA6D /PLLP/ TMCC3	6	0.199	0.017	Caudate basal ganglia	ensheathment of neurons, axon ensheathment, myelination, neurogenesis,

					oligodendrocyte differentiation, glial cell differentiation, nervous system development, generation of neurons
GPATCH2*/NCAM2/ERBB4/RPAP3/ SACM1L*	5	0.224	0.022	Brain putamen basal ganglia	mitochondrial ATP synthesis coupled electron transport, ATP synthesis coupled electron transport, aerobic electron transport chain, respiratory electron transport chain, oxidative phosphorylation
UNC50/RPAP3/SACM1L	3	0.327	0.025	Brain cervical spinal cord	sterol biosynthetic process, secondary alcohol biosynthetic process, cholesterol biosynthetic process, secondary alcohol metabolic process, alcohol biosynthetic process, cholesterol metabolic process, central nervous system development sterol metabolic process
GPATCH2/***TLE3****/CDC42BPA*	3	0.345	0.049	Brain substantia nigra	regulation of histamine secretion by mast cell
SSUH2*/RPAP3/ZNF496/SLC48A1/ DCTN4/WDR27*	6	0.367	0.049	Amygdala	nitrogen compound metabolic process, cellular metabolic process
QDPR/DTX4/CDC42BPA/PLLP /TMCC3	5	0.415	0.048	Brain nucleus accumbens basal ganglia	ensheathment of neurons, axon ensheathment, myelination, neurogenesis, oligodendrocyte differentiation, glial cell differentiation, nervous system development, generation of neurons

Table S7. Module enrichment analysis performed using eMAGMA. The Genes column indicates the gene set enriched in our Add Health and ISP meta-analysis and gene count indicates the number of genes in each module. Genes highlighted in bold were identified as a nearest gene for a meta-analysis top hit (see Table 2). Module enrichment reported for any gene tissue-specific analysis with a raw $p < 0.05$, "Adj pval" indicates the FDR corrected p value. A competitive gene pathway analysis was performed on tissue-specific gene co-expression modules using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost). The biological process column describes biological pathways of tissue-specific modules enriched with stuttering gene-based signals.

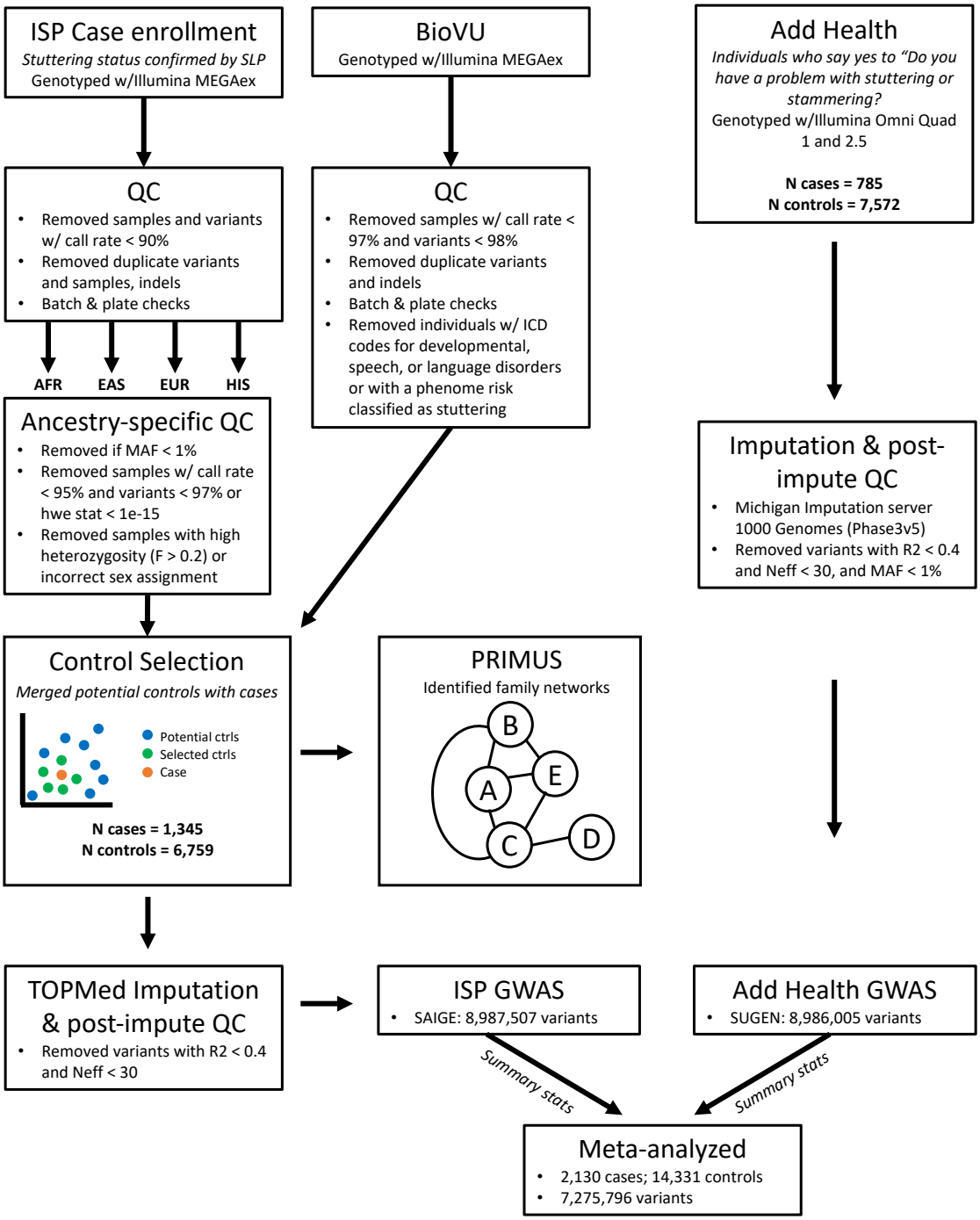

Figure S1. Experimental workflow. Diagram indicates the flow and sequence of applied quality control filters for each analyzed dataset (ISP cases and controls and Add Health samples). Details for each step are depicted in the methods section.

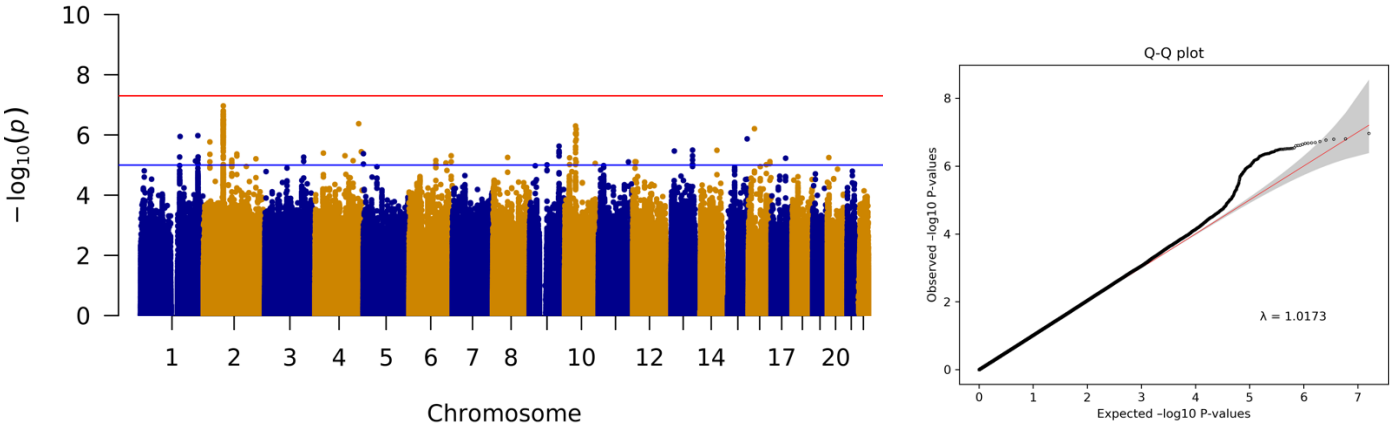

Figure S2. Manhattan and qq-plot of stuttering genome-wide association results in ISP. Analysis included 1,345 cases and 6,759 ancestry and sex-matched controls for 8,987,507 variants imputed across autosomes using the TOPMed reference. No loci reached genome-wide significance (red line $p < 5 \times 10^{-8}$). Nineteen loci reached suggestive genome-wide significance (blue line $p < 5 \times 10^{-6}$).

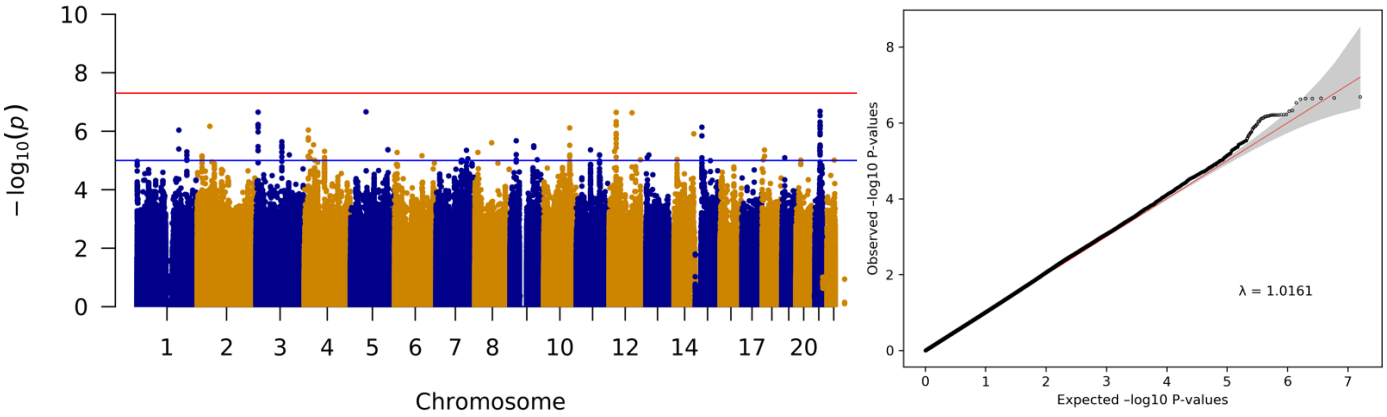

Figure S3. Manhattan and qq-plot of stuttering genome-wide association results in Add Health. Analysis included 785 cases and 7,572 controls for 8,986,005 variants imputed across autosomes using the 1000 Genomes Reference. No loci reached genome-wide significance (red line $p < 5 \times 10^{-8}$); twenty-four loci reached suggestive genome-wide significance (blue line $p < 5 \times 10^{-6}$).

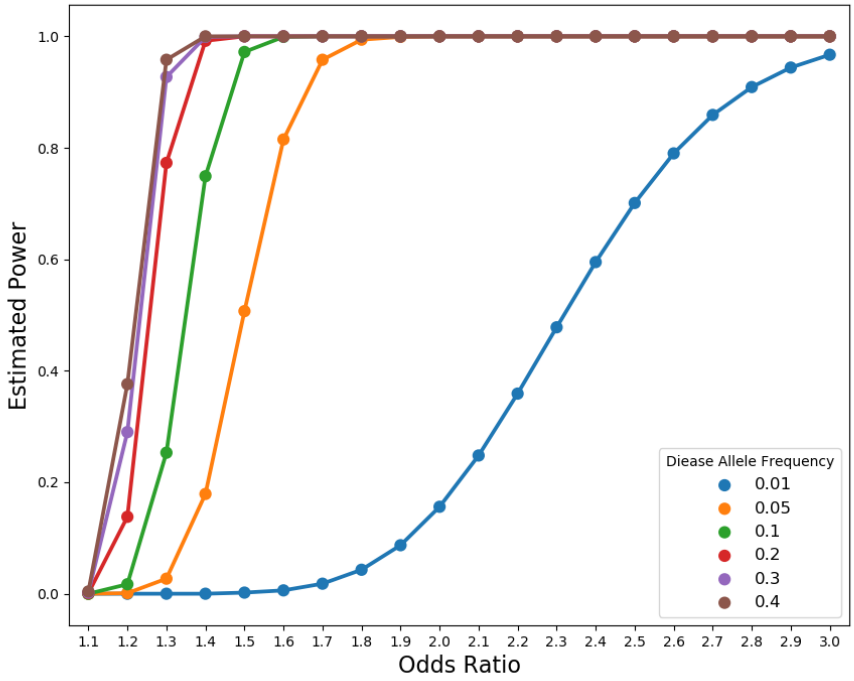

Figure S4. Power to detect significant developmental stuttering associations in meta-analysis. Power calculation included 2,130 cases and 14,331 controls and assumed a two-sided hypothesis test at $p < 5 \times 10^{-8}$, an additive model, and a developmental stuttering population prevalence of 1%. Each color represents estimated power for the disease allele frequency indicated in the figure legend.