

Identifying novel putative genes linked to stuttering in highly multiplex Indian families using exome sequencing

SRIKUMARI CR¹, NANDHINI G¹, and Chandru J¹

¹University of Madras

October 8, 2020

Abstract

Abstract Stuttering is a complex speech disorder and heritability of this trait is persuasive, with multiple afflicted families showing phenotypic segregation across generations, yet no conclusive genetic etiology could be identified. Analyzing multiplex families using exome sequencing(ES) may help in identification of putative genes and scope for understanding the mutational burden for speech implicated pathways. In this study ES was performed in six individuals from two clinically well characterized, multiple affected, south Indian families (STU-65 and STU-66) showing stuttering across five generations. From ES to variant prioritization, a sequential bioinformatics approach was implemented to search for putative gene targets. In the two multiplex families studied, ES data analysis resulted in an enriched list of 14 genes (with variants) (COL4A2,COL6A3,COL6A6,ITGAX,LAMA5,ADAMTS9,CSGALNACT1, TMOD2,HTR2B,RSC1A1,TRPV2,WNK1,ARSD and SPTBN5) involved in neural functions. Additionally, a homozygous variant in NLRP11 gene and a heterozygous variant in NAGPA gene were identified in STU-65 family that needs further confirmation. Our results support the fact that stuttering is a polygenic disorder. The putative gene targets identified in our study can drive the research prospects to understand the underlying mechanisms. We hypothesize multiple and combined mechanisms to be involved in the genesis of stuttering. Keywords: Stuttering, exome sequencing, neural pathways

INTRODUCTION

Stuttering is a disorder of fluency that disrupts the normal flow of speech wherein people who stutter (PWS) know what they want to say but have trouble saying it. The condition is generally outgrown by children but it persists in some as an enduring problem. Emergence of stuttering is multifactorial that includes genetic, environmental and neurological factors. Environmental factors though not a cause, can interact with genes (cause epigenetic modifications) and may aggravate the dysfluency under conditions of stress, anxiety and also depends on family dynamics, child's temperament, etc¹.

The evidence for genetic factors were indicated in multiple studies worldwide². Worldwide search for causative genes began with linkage studies, as applicable to any inherited disorder, to elucidate the genetic and non-genetic etiology relevant to stuttering. Nevertheless since stuttering is a complex trait, linkage analysis resulted in mapping multiple loci across studies. Identification of genes was not possible until studies focused on consanguineous families³. The causative genes linked to stuttering were identified to be *GNPTAB*, *GNPTG*, *NAGPA* and *AP4E1* that explains 20% of stuttering cases⁴. All these genes point to intracellular trafficking deficits, an emerging concept in neurological disorders, implicating lysosomal dysfunction but the mechanism linking these genes to the phenomena of stuttering is unknown. However these deficits are said to have unique impact on specific neurons responsible for speech. A recent mouse model study tried to illuminate the neurons that are central to the mystery of stuttering and indicated that abnormality lies in astrocytes (essential for nerve activity) of corpus callosum (largest white matter structure). This limited astrocytes population slows down inter-hemispheric communication by a tiny bit, noticeable only in speech⁵.

Speech production requires efficient communication between central nervous system (cortical and sub-cortical regions) and peripheral nervous system (cranial nerves that innervate respiratory, phonatory, articulatory and resonance systems). Age of onset of stuttering parallels the developments in anatomy of speech system, language and articulatory skills. Interferences in these maturational processes may contribute to stuttering, hence brain imaging studies in children could enhance our understanding of pathophysiology⁶.

Brain imaging studies revealed structural and functional connectivity differences in Children Who Stutter (CWS) involving neural networks that support self-initiated timing of speech movements (BGTC - Basal Ganglia Thalamo-cortical loop, SMA and Putamen) and in integration of auditory feedback to speech motor control process. Structural (white matter) connectivity involving corpus callosum and those among the putamen and cortical motor and auditory regions is attenuated in the left hemisphere in children who stutter^{7,8}. Over activity of the dopamine neurotransmitter has been implicated as a potential factor in developmental stuttering⁹. Thus it is evident that the problem is not confined to a single part of brain rather connecting different parts of the brain.

The findings in stuttering research - overactive dopamine, deficits in neural circuits and genes involved in cellular waste disposal - all point to communication deficits. All these findings hint that stuttering may result from slight delay in communication with different parts of the brain.

Research studies on the genetics of stuttering in India are limited and nearly unexplored. Our preliminary genetic epidemiology study in a large cohort of 74,544 school children in the state of Tamil Nadu, reported the prevalence of stuttering to be **0.46%**¹⁰. Investigational mutation screening for the recurrence of previously reported stuttering genes showed a minimal resolution of 3.1% (2/64) that could be ascribed to these genes but remains elusive, which is reported as the first paper in this series.

Since stuttering is a complex trait that segregates in families, analyzing multiplex families would help in either identification of highly penetrant rare variants or variants that are common in affected individuals. Subsequently this forms the concept of this paper as the second in this series. ES was employed in two multiple affected families from existing database¹⁰. The aim of the study is to identify new candidate genes for stuttering, by analyzing the pathogenic variants inherited in the multiplex families but not present in public databases.

RESULTS

ES is a powerful technology to reveal the complex genetic architecture and unravel the missing heritability. ES was performed in six individuals from two highly multiplex families with severe stuttering: STU-66 nuclear family (figure 1) comprising of affected father (IV-2), unaffected mother (IV-3), unaffected brother (V-4) and the proband (V-5); STU-65 family (figure 2) includes two affected siblings (V-33;V-35). Further details of the families are mentioned in the methodology section.

Variant filtering

STU-66 family: ES data of the four members, resulted in 73,426 variants in affected father, 73,689 in unaffected mother, 73,141 in unaffected sibling and 73,198 variants in proband.

Applying variant filter criteria (detailed in methodology), resulted in a mean of 2245 variants (ranging from 2188-2296). The variant data of the four individuals were analyzed using Venn diagram that produced 15 possible combinations of the data (figure 3). One such combination revealed 276 genes that were common to affected father and the proband (figure A1). On filtering further for the common variants in them ensued 260 variants harbored in 248 genes. Of these 260 common variants, 218 were previously known (dbSNP database) and 42 were novel variants (hitherto unreported). Only 85 of the 218 known variants and 12 of the 42 novel variants fall in the coding region. Applying a four step exclusion criteria (detailed in methodology) to known variants, shortened the list to 18 genes (all heterozygous variants). The genes of interest thus sums up to 19 known variants and 12 novel variants, spread across 30 genes (table A1).

STU-65 family: ES data of the two affected sibs (V-33, proband and V-35, affected brother) resulted in

58,346 and 51,062 variants respectively. Applying similar criteria as described above using Venn diagram approach, 294 genes were found common to the sib-pair (figure A2). The downstream analysis of the variants observed in the 294 common genes is depicted in the flow chart (figure A3). Those variants common to both sibs were sorted out resulting in 161 common variants harbored in 153 genes. Of these, 158 were known variants and 3 were novel variants that were all found in the coding regions. Applying the same four step exclusion criteria (as in STU-66) to 158 known variants, narrowed down the gene list to 34 variants (4 homozygous and 30 heterozygous). Thus the genes of interest sum up to 31 known and 3 novel variants, in 34 genes. (table A2).

No significant intronic/UTR variation was detected, other than the indels and missense variations, in the exome data of both families. Considering stuttering to be a less common trait, an allele count of more than 20 in gnomAD database were excluded, narrowing the list to 35 genes.

Comparison of variants with available databases

In the STU-66 family, only one missense variant in *COL4A2* gene (c.461G>A; p.Gly154Glu) previously reported in a disorder related to fluency of language¹¹ was observed in our study. *COL4A2* encodes type IV collagen that forms a major structural component of basement membrane and are involved in cerebrovascular disease that include porencephaly and white matter lesions^{12,13}.

Since the STU-65 family had consanguinity for three generations, we scrutinized homozygous variants first. Only one *NLRP11* variant (c.1052G>A; p.Arg351Gln) was present in homozygous condition in both the sibs. *NLRP11* is a member of NOD like receptor protein with pyrin domain responsible for formation of inflammasome but it is said to represses NF- κ B and type I interferon responses, involved in inflammation pathways¹⁴. Neuro-immune and neuro-inflammation has been established as key factors in the development of Autism, where speech is also affected¹⁵. *NLRP11* could be a possible candidate gene in this family. In addition a recently reported heterozygous variant in *NAGPA* gene (c.322G>A; p.Glu108Lys; gnomAD database) implicated in stuttering was also observed in sibs.

Identification of genes in common pathways using systems biology approach

Checking manually for the phenotypic relevance of the 64 genes yielded nine putative pathways encompassing 41 genes. This can be directly or indirectly linked to stuttering (table 1). Intriguingly, the possible candidate genes are different for the two families, but they fit into nine common pathways. Similarly, pathway browser tool in Reactome database (<https://reactome.org>) determine pathways that are over-represented. In this manner 25 most relevant pathways encompassing 14 genes (table 2), were sorted based on their p-value and the remaining genes were indicated as interactors (table A3). Thus the enriched gene list comprised of 14 genes that recurred in the two approaches (tables 1 and 2).

Genes commonly observed in the two multiplex stuttering families

To determine the genes causing stuttering, the expansive list of genes with variations, that was exclusive to the affected individuals in each family was compared (276 genes in STU-66 family and 294 in STU-65). Seven genes (*SLC36A1*, *C1orf167*, *SYNC*, *TM4SF1*, *MUC6*, *SPTBN5* and *MSRB1*) were found to be common to both families, irrespective of the variants identified (figure A4). All the affected members harbored multiple variants in each of these seven genes.

DISCUSSION

The findings in exome sequencing of two multiplex families with severe stuttering are discussed. Considering the previously implicated genes (*GNPTAB*, *GNPTG*, *NAGPA* and *AP4E1*) only one pathogenic heterozygous missense variant (c.322G>A; p.Glu108Lys) in *NAGPA* was detected in our study. Since the so far implicated genes in stuttering are limited to a single lysosomal pathway, we extended the analyses to the 123 genes also, known in lysosomal function, but did not identify any significant variation in them in our two families.

Search for other putative pathways

A segregation based strategy was used to account for those gene variants overlapping with the affected members (276 genes in STU-66 family and 294 in STU-65) but absent in the unaffected members. Seven genes viz., *SLC36A1*, *C1orf167*, *SYNC*, *TM4SF1*, *MUC6*, *SPTBN5* and *MSRB1* were common to both the families with stuttering though they had different sets of multiple variants in each gene. *SLC36A1* encodes a proton dependent amino acid transporter, highly expressed in brain. Mutations in this gene involves poor coordination of speech associated with iminoglycinuria and spinocerebellar ataxia 45. Mutations in *C1orf167* is associated with neural tube defects. *SYNC* gene is involved in muscle contraction and is regarded as a marker of neuromuscular disease. *TM4SF1* and *MUC6* genes form the extracellular components. *SPTBN5* gene encodes protein that links membrane lipids, proteins and cytosolic factors to cytoskeletal filaments and also binds to kinesin and actin proteins. *MSRB1* gene acts in response to stress and is also involved in actin repolymerization (<https://www.genecards.org>). Analysis of protein-protein interactions (<https://string-db.org/>) of the protein products of these seven genes using *in silico tool* (figure A5), the gene products did not show any interaction with each other but they seem to be involved in a common biological mechanism: within cell or cell-cell communication. These deficits at neuromuscular junctions might contribute to stuttering as speech involves 40,000 neuromuscular events per second that act in coordination¹⁶.

In each family the downstream analysis of genes that overlap only among affected members resulted in a set of 30/276 genes in STU-66 and 34/294 genes in STU-65. These 64 genes when subjected to pathway browser tool (that interact with Reactome biological pathways), to look for common biological mechanisms, resulted in an enriched gene list of 14 target genes (*COL4A2*, *COL6A3*, *COL6A6*, *ITGAX*, *LAMA5*, *ADAMTS9*, *CSGALNACT1*, *TMOD2*, *HTR2B*, *RSC1A1*, *TRPV2*, *WNK1*, *ARSD* and *SPTBN5*) with P-values ranging from 0.125-0.0001 (figure 4). Since the genes that interact significantly tend to have a common function and may be involved in the same pathway, functional dissection of 14 possible candidate genes was attempted to identify the pathways related to stuttering.

Genes like *COL4A2*, *COL6A3*, *COL6A6* are involved in formation and degradation of collagens¹⁷, *ITGAX* for integrin¹⁸ and *LAMA5* for laminin¹⁹ that together form the extra cellular matrix (ECM) components. *ADAMTS9* gene codes for metalloproteinase involved in ECM degradation²⁰. *CSGALNACT1* is involved in chondroitin sulfate biosynthesis, one of the GAGs associated to core protein as proteoglycans that also form the component of ECM²¹.

ECM has three basic parts that involves basal lamina, interstitial matrix and perineuronal networks. ECM is dynamic, provides structural stability in the form of basement membrane, plays role in signaling and are critical for neuroplasticity^{22,23}. ECM provides an appropriate environment for the development and functioning of cells, like muscle and nerve cells^{24,25}. In the present study, variations in genes encoding collagen and laminins may disrupt basement membrane which has important role in neuroplasticity and maintaining synapses.

Other gene variants observed in the present study are in *TMOD2* gene involved in muscle contraction, *HTR2B* in signaling serotonin receptors, *RSC1A1*, *TRPV2* and *WNK1* genes in transport of small molecules in stimuli-sensing channels. All of them are involved in communication between neuron and muscle cells and may coordinate production of speech. At neuro muscular junctions (NMJ), motor neurons release neurotransmitters that bind to post-synaptic receptors, leading to contraction of muscles²⁶. Hence disturbances in signaling system might occur at NMJ that may disrupt speech.

ARSD belongs to cluster of sulfatase genes that hydrolyse GAGs that are found to represent truncated pseudogenes²⁷. *SPTBN5* is involved in vesicle transport and NCAM signaling (neural cell adhesion molecule) that plays an important role in nervous system development and synaptic plasticity²⁸.

Thus most of the genes identified form extra cellular matrix molecules, involved in signaling functions indicating overall cell communication deficits. These pathways very well coincide with findings of neuroimaging (deficits in white matter tracts)^{8,29} and genetic animal studies (deficits in corpus callosum)⁵ which conclude cell communication deficits as the major cause for stuttering. Shugart *et al.*,³⁰ implicated, interesting candidate genes (on chromosome 18) like desmoglein/desmocolin family and neuronal cadherin 2 gene that helps

in cell-cell communication and cell adhesion. Such communications are important in neurons involved in the production of speech. As opposed to traditional thinking that origin of speech and language could be localized to one part of our brain, today neural bases are in fact found in the circuits that connect different regions of brain³¹. In a more elaborate GWAS study, 10 candidate genes were identified to be strongly associated with stuttering³². Globally the genes implicated in stuttering till date, in all affected individuals were reported to have only one mutant copy with reduced level of functional protein (haploinsufficiency) rather than total absence³³.

Our ES approach in two multiplex stuttering families point to interesting putative genes involved in signaling and transport. However in STU-65 family, a homozygous variant in *NLRP11* gene and a heterozygous variant in *NAGPA* gene were also identified. The role of these variants must be explored by confirming its segregation in the other affected members of the family and also through functional validation.

Our investigation is of a descriptive type and the potential candidate genes identified should be viewed with caution because some of the gene variants may not have direct role in the development of the disorder. ES in additional members would further narrow down the variants. Functional studies of the identified variants would help in understanding the biological mechanism. Yet, our results support the fact that stuttering is a polygenic disorder with a complex genetic background and we hypothesize multiple and combined mechanisms to be involved in the genesis of stuttering. The variable expressivity of the disorder in our families could be explained by this genetic heterogeneity.

ES addresses only protein coding regions of genes within genome (1.5%); 80% of DNA that was believed to be junk is now biologically active that makes RNA, of which 20% make functional RNAs involved in regulation (<http://jonlieffmd.com/blog/mind-and-molecular-genetics-in-the-neuron-3-evolution>). The regulatory regions are 20 times bigger than protein coding regions. Thus the complexity lies not only in the number of genes involved but also may be due to the huge regulatory dynamics.

CONCLUSION

Dissecting the genetic architecture will not only contribute to a better understanding of the etiology of stuttering but also facilitate the identification of new targets with implications to prevention, diagnosis and treatment. Using ES, we generated an enriched list of 14 target genes (*COL4A2*, *COL6A3*, *COL6A6*, *ITGAX*, *LAMA5*, *ADAMTS9*, *CSGALNACT1*, *TMOD2*, *HTR2B*, *RSC1A1*, *TRPV2*, *WNK1*, *ARSD* and *SPTBN5*) involved in neural related pathways, in the two multiplex families studied. We propose that these 14 genes having role in signaling and transport may be putatively linked to stuttering. However in STU 65 family, a homozygous variant in *NLRP11* gene and a heterozygous variant in *NAGPA* gene were also identified. ES in additional members and further functional validation would ensure the role of these genes in stuttering.

METHODOLOGY

Two multiplex families (STU-65 and STU-66) were recruited from database and clinically characterized for stuttering by speech pathologist using Stuttering Severity Instrument-3³⁴. Using a structured questionnaire family history and other details were noted. Eight milliliters of blood was collected from probands and their family members. Genomic DNA was isolated using PCI extraction method³⁵ and quantified by Nano drop (Thermo- Fisher Scientific, Wilmington, USA). About 500ng of DNA per sample was quantified and 20 μ L volume of each sample was sent to Medgenome Labs Ltd., Bengaluru for comprehensive exome sequencing. ES was performed in six individuals; four individuals from STU-66 family and two affected siblings from STU-65 family.

STU-66 family: A Telugu speaking family showed seven affected across three generations with three inbreeding loops in the pedigree (figure 1). The proband (V-5) was born to non-consanguineous parents, where the father (IV-2) had mild stuttering but mother (IV-3) was unaffected. Both the proband (V-5) and his younger brother (V-6) had moderate but progressive stuttering, with age at onset of 2.5 years; his elder brother (V-4) was unaffected. His grandmother (recovered; III-1), uncle (mild stuttering; IV-7) and two first

cousins (severe; V-2,V-3) were also affected with stuttering. The nuclear family comprising of affected father (IV-2), unaffected mother (IV-3), unaffected brother (V-4) and the proband (V-5) were selected for ES.

STU-65 family: A Tamil speaking family from a major endogamous Mudaliar caste had more than 20 PWS across six generations. Availability of senior informants starting from third generation helped to trace intense multigenerational inbreeding, tracing the phenotype to a common female founder I-2 (figure 2).

The proband (V-33) and his affected brother (V-35) were born to consanguineous parents (III-21,IV-8) who were also affected with stuttering. Though his mother complained of severe stuttering in the early days, her dysfluencies have reduced now but with a reminiscence of fast speech rate and jaw clenching. Father also had repetitions and jaw clenching.

Both the proband (V-33) and his younger brother (V-35) developed stuttering gradually at 2.5 years and were moderate and severe respectively with no birth complications. The proband was a dropout from school and had situational increase, with strangers. Dysfluencies observed include hard contacts in initial syllable, prolongation, silent pauses, syllable and part-word repetitions with iterations of 2-3. Secondary behaviours included eye blinks, clicking sounds, fixed articulatory posture, nose flaring, tension in the neck, jaw jerking and frequent left side head nod. His rate of speech was slow and intelligibility in speech was fair. His brother had repetitions and prolongations along with eye blinking, facial grimace, hand fidgeting, etc. He also had situational increase but continued beyond school education. Their grandmother (mild), maternal aunts and cousins were also affected with severe stuttering. The extended family members have been extensively phenotyped (table A4).

ES was commercially carried out at MedGenome Labs Ltd., Bangalore facility. ES library was prepared using Agilent-Sure Select XT Reagent Kit (Illumina). Biotinylated oligonucleotide capture probes (V5+UTR) designed for all the coding exons were used to enrich the region of interest (exome) by hybridization. The library obtained was diluted to final concentration of 2nm in 10ul and subjected to Cluster amplification. The flow cell was loaded on to the sequencer (Hi Seq X10) to generate 2X150 bp sequence reads at 100x sequencing depth. Sequenced data with Q30 values was considered as qualified and processed to generate FASTQ files for further downstream variant analysis.

The raw data was quality trimmed and reads (using fastq-mcf command line tool) were aligned to Human Reference genome (alignment to hg19) using BWA-MEM tool. The output in SAM format is converted to BAM file (using Samtools) and processed to obtain SNVs (Single Nucleotide Variation) and INDELS (small insertions and deletions) in a standard VCF (Variant call format) file. Coverage of the genes were analyzed using Bedtools.

The variants were called using GATK software and annotated using MedGenome in-house variant annotation pipeline (VariMAT - Variation and Mutation Annotation Toolkit). It integrates multiple clinical grade databases [GWAS (<https://www.genome.gov/genetics-glossary/Genome-Wide-Association-Studies>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>), OMIM (<https://www.omim.org>), UniProt (<https://www.uniprot.org/>), ExAC (<http://exac.broadinstitute.org>; <https://gnomad.broadinstitute.org>), dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>), 1000 genomes (<https://www.coriell.org/1/NHGRI/Collections/1000-Genomes-Collections/1000-Genomes-Project>)], variant class prediction and variants pathogenicity prediction tools for annotating the variants. VariMAT annotated variants contain information on the population frequency, computational pathogenicity prediction, variant type and predicted impact of the variant on the protein (missense, loss of function, etc.).

Sequencing data of STU-66 family: The paired-end ES, generated a data of 10-14 Gb for each of the four individuals sequenced. More than 93% of the data showed variant quality scores above Q30 (Q score measures the base calling accuracy by estimating base calling error probabilities and Q30 indicates the probability of incorrect base call of 1 in 1000 which means the base call accuracy is 99.9%). The overall alignment and the passed alignment percentage in all the samples was around 99.99 and 97.28 % respectively. The analysis was performed after alignment using SS-V5-UTR panel (74,557,381bp) which covers 23690 genes. The average panel depth for each of the sample ranged from 80.42 to 99.27X.

Sequencing data of STU-65 family: In this family the proband and his affected brother, V-33 and V-35, were sequenced generating total data of 8-11 Gb. More than 90% of the data showed quality score distribution above Q30. The overall alignment and passed alignment percentage (alignment to hg19) in proband and sib was around 99.99 and 94.89 % respectively. The average panel depth for each sample ranges from 80 to 100 X.

Variant filtering: The downstream variant analysis protocol involves short listing of the variants in the affected individuals, followed by prioritization of variants based on the data from additional family members. Each of the individual variant files were subjected to variant filtering criteria, to profile *rare variants* in all samples under investigation, comprising of:

1. The high impact variant type that includes frameshift, termination, start loss type,
2. The moderate impact variants being missense, stop loss and indels, and
3. Intronic splice site variants comprise of splice donor/acceptor or proximal splicing impact variants. These variants are filtered with a cutoff for depth being 3X and MAF <1% in ExAC, 1000 genomes, and proprietary Medvar (MedGenome variation) databases.

Variant prioritization: The variant prioritization and candidate gene identification in family based cohort often rely on profiling the common and uncommon gene variants present in the affected and unaffected members in a family. An open source tool called InteractiVenn³⁵ was deployed for this purpose that can handle up to six data sets (creates Venn diagram). To this resulting variant list the following exclusion criteria were employed.

Exclusion criteria

1. those predicted as tolerant using MetaSVM (meta-analytic support vector machine (SVM)) and MetaLR which accommodate multiple omics data to detect consensus genes associated with disease, were excluded. In other words, those variants predicted to be damaging by more than one prediction tool was considered.
2. In sorting the known phenotype, variants in gene with completely unmatched OMIM phenotypes were excluded.
3. Proximal splice site variants placed at >8 nucleotides away from exon were also excluded.
4. Variants in zinc finger encoding genes were excluded as they have very omnipresent role.

Comparison of variants with available databases

These shortlisted variants were matched with previously reported genes, attributed to stuttering or any other speech and language disorders

1. Significant SNPs reported by Kraft³² in the 'Genome-wide association study of persistent developmental stuttering' of northern European ancestry.
2. The 123 genes listed in the lysosomal pathway (<https://pathcards.genecards.org/card/lysosome>). Since the three genes so far known to be implicated in stuttering were identified in the lysosomal pathway, this study carefully observed any potential variants in genes related to this pathway.
3. The genes so far listed for speech and language disorders by³⁷.

Bibliography

1. Anderson JD, Pellowski MW, Conture EG KE. Temperamental characteristics of young children who stutter. J Speech, Lang Hear Res. 2003;
2. Kraft SJ, Yairi E. Genetic bases of stuttering: The state of the art, 2011. Vol. 64, Folia Phoniatica et Logopaedica. 2011. p. 34-47.
3. Frigerio-Domingues C, Drayna D. Genetic contributions to stuttering: the current evidence. Mol Genet Genomic Med [Internet]. 2017 Mar;5(2):95-102. Available from: <http://doi.wiley.com/10.1002/mgg3.276>

4. Raza MH, Domingues CEF, Webster R, Sainz E, Paris E, Rahn R, et al. Mucopolidosis types II and III and non-syndromic stuttering are associated with different variants in the same genes. *Eur J Hum Genet.* 2016;
5. Han, T. U., Root, J., Reyes, L. D., Huchinson, E. B., du Hoffmann, J., Lee, W. S., ... & Drayna D. Human GNPTAB stuttering mutations engineered into mice cause vocalization deficits and astrocyte pathology in the corpus callosum. *Proc Natl Acad Sci.* 2019;116(35):17515–24.
6. Watkins RV, Yairi E AN. Early childhood stuttering III: Initial status of expressive language abilities. *J Speech, Lang Hear Res.* 1999;Oct,42(5):1125–35.
7. Chang SE, Zhu DC. Neural network connectivity differences in children who stutter. *Brain.* 2013;136(12):3709–26.
8. Chang, S. E., Zhu, D. C., Choo, A. L., & Angstadt M. White matter neuroanatomical differences in young children who stutter. *Brain.* 2015;138(3):694–711.
9. Comings DE, Wu S, Chiu C, Ring RH, Gade R, Ahn C, MacMurray JP, Dietz G MD. Polygenic inheritance of Tourette syndrome, stuttering, attention deficit hyperactivity, conduct, and oppositional defiant disorder: The additive and subtractive effect of the three dopaminergic genes—DRD2, D β H, and DAT1. *Am J Med Genet.* 1996;May 31;67(3):264–88.
10. Nandhini Devi G, Thalamuthu A, Valarmathi S, Karthikeyen NP, Srikumari Srisailapathy CR. Genetic epidemiology of stuttering among school children in the state of Tamil Nadu, India. *J Fluency Disord.* 2018;58.
11. Eicher, J. D., Powers, N. R., Miller, L. L., Akshoomoff, N., Amaral, D. G., Bloss, C. S., ... & Chang L. Genome-wide association study of shared components of reading disability and language impairment. *Genes, Brain Behav.* 2013;12(8):792–801.
12. Verbeek, E., Meuwissen, M. E., Verheijen, F. W., Govaert, P. P., Licht, D. J., Kuo, D. S., ... & Halley DJ. COL4A2 mutation associated with familial porencephaly and small-vessel disease. *Eur J Hum Genet.* 2012;20(8):844.
13. Yoneda, Y., Haginoya, K., Arai, H., Yamaoka, S., Tsurusaki, Y., Doi, H., ... & Matsumoto N. De novo and inherited mutations in COL4A2, encoding the type IV collagen α 2 chain cause porencephaly. *Am J Hum Genet.* 2012;90(1):86–90.
14. Ellwanger, K., Becker, E., Kienes, I., Sowa, A., Postma, Y., Gloria, Y. C., ... & Kufer TA. The NLR family pyrin domain-containing 11 protein contributes to the regulation of inflammatory signaling. *J Biol Chem.* 2018;293(8):2701–10.
15. Siniscalco, D., Schultz, S., Brigida, A. L., & Antonucci N. Inflammation and neuro-immune dysregulations in autism spectrum disorders. *Pharmaceuticals.* 2018;11(2):56.
16. Logan R. The Three Dimensions of Stuttering-Neurology, Behaviour, and Emotion. *J Fluency Disord.* 2000;4(25):387–388.
17. Haq, F., Ahmed, N., & Qasim M. Comparative genomic analysis of collagen gene diversity. *3 Biotech.* 2019;9(3):83.
18. Takada, Y., Ye, X., & Simon S. The integrins. *Genome Biol.* 2007;8(5):215.
19. Maselli, R. A., Arredondo, J., Vázquez, J., Chong, J. X., Bamshad, M. J., Nickerson, D. A., ... & McDonald CM. A presynaptic congenital myasthenic syndrome attributed to a homozygous sequence variant in LAMA5. *Ann N Y Acad Sci.* 2018;1413(1):119–25.
20. Kelwick, R., Desanlis, I., Wheeler, G. N., & Edwards DR. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol.* 2015;16(1):113.

21. Uyama, T., Kitagawa, H., Tanaka, J., Tamura, J. I., Ogawa, T., & Sugahara K. Molecular cloning and expression of a second chondroitin N-acetylgalactosaminyltransferase involved in the initiation and elongation of chondroitin/dermatan sulfate. *J Biol Chem.* 2003;278(5):3072–8.
22. Hynes. The extracellular matrix: Not just pretty fibrils. *Science* (80-). 2009;326(5957):1216–9.
23. Long, K. R., & Huttner WB. How the extracellular matrix shapes neural development. *R Soc Open Biol.* 2019;9(1):180216.
24. Barros, C. S., Franco, S. J., & Müller U. Extracellular matrix: functions in the nervous system. *Cold Spring Harb Perspect Biol.* 2011;3(1):a005108.
25. Grzelkowska-Kowalczyk K. The importance of extracellular matrix in skeletal muscle development and function. *Compos Funct Extracell Matrix Hum Bodyq.* 2016;3-24.
26. Wu, H., Xiong, W. C., & Mei L. To build a synapse: signaling pathways in neuromuscular junction assembly. *Development.* 2010;137(7):1017–33.
27. Meroni, G., Franco, B., Archidiacono, N., Messali, S., Andolfi, G., Rocchi, M., & Ballabio A. Characterization of a cluster of sulfatase genes on Xp22. 3 suggests gene duplications in an ancestral pseudoautosomal region. *Hum Mol Genet.* 1996;5(4):423–31.
28. Kleene, R., Mzoughi, M., Joshi, G., Kalus, I., Bormann, U., Schulze, C., ... & Schachner M. NCAM-induced neurite outgrowth depends on binding of calmodulin to NCAM and on nuclear import of NCAM and fak fragments. *J Neurosci.* 2010;30(32):10784–98.
29. Kronfeld-Duenias, V., Civier, O., Amir, O., Ezrati-Vinacour, R., & Ben-Shachar M. White matter pathways in persistent developmental stuttering: Lessons from tractography. *J Fluency Disord.* 2018;55:68–83.
30. Shugart YY, Mundorff J, Kilshaw J, Doheny K, Doan B, Wanyee J, et al. Results of a genome-wide linkage scan for stuttering. *Am J Med Genet [Internet].* 2004 Jan 15;124A(2):133–5. Available from: <http://doi.wiley.com/10.1002/ajmg.a.20347>
31. Lieberman, P., & McCarthy R. Tracking the evolution of language and speech-Comparing Vocal Tracts to Identify Speech Capabilities. *Exped Mag Univ Pennsylvania.* 2007;49(2):15–20.
32. Kraft SJ. Genome-wide association study of persistent developmental stuttering. University of Illinois; 2010.
33. Kang C, Drayna D. A role for inherited metabolic deficits in persistent developmental stuttering. Vol. 107, *Molecular Genetics and Metabolism.* 2012. p. 276–80.
34. Riley GD. Stuttering Severity Instrument-3 (SSI-3). 1994;Austin, TX.
35. Sambrook, J.F. and Russell D. *Molecular Cloning: A Laboratory Manual.* 3rd ed., V. Press CSHL, editor. 2001.
36. Heberle, H.; Meirelles, G. V.; da Silva, F. R.; Telles, G. P.; Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics.* 2015;16:169.
37. Guerra J CR. Genomics of speech and language disorders. *Transl Genet Genom.* 2019;3(9).

Table 1: Manually curated list of candidate* genes identified in 9 putative pathways, that can be directly or indirectly linked to stuttering, in the two families studied

S.No	Gene list from STU-66 family	Pathways	Gene list from STU-65 family
1	SPTBN5, <i>PLXNA1</i> , <i>RAP1GAP</i>	neural developmental pathways	<i>COL6A6</i> , <i>NELL2</i> , <i>DSCAML1</i>

S.No	Gene list from STU-66 family	Pathways	Gene list from STU-65 family
2	ADAMTS9, <i>COL6A3</i> , <i>COL4A2</i>	extra cellular matrix organisation/degradation	LAMA5 , <i>COL6A6</i> , ITGAX
3	<i>TMOD2</i> , <i>SYNC</i>	muscle contraction	MYHI
4	CD3EAP	gene expression	-
5	SPTBN5, <i>ZNF763</i> , <i>CCNJL</i> , <i>KIF13B</i> , <i>ATXN2</i>	signal transduction/ vesicle transport	<i>ATXN2L</i> , <i>HTR2B</i> , <i>STARD13</i> , <i>MYH1</i> LAMA5 , <i>NEK4</i> , <i>MRIP</i> , <i>KDM5A</i> TRPV2, WNK1
6	<i>RSC1A1</i> , <i>NIPA2</i>	transport of small molecules	<i>PAN3</i> , <i>ARSD</i> , <i>WDR27</i> , <i>HSIT3H2A</i>
7	<i>CSGALNACT1</i> , ADAMTS9, SPTBN5, <i>ALAS1</i>	metabolism	<i>CRTAM</i> , ITGAX , NLRP11
8	<i>GPR97</i> , SPTBN5, <i>ABCA13</i> , <i>UNKL</i>	immune system	<i>NAGPA</i>
9	-	Lysosome enzyme targeting	

Note: *The candidate genes proposed are exclusive to affected individuals, segregating in heterozygous status, except for NLRP11 segregating in the homozygous state; **bold** indicates genes involved in multiple pathways

Table 2: Tentative list of candidate* genes identified in enrichment pathways, that can be directly or indirectly linked to stuttering, in the two families studied

S.No	Gene list from STU-66 family	Pathways	Gene list from STU-65 family
1	<i>COL4A2</i> , <i>COL6A3</i>	ECM proteoglycans	<i>COL6A6</i> , ITGAX , <i>LAMA5</i>
2	<i>COL4A2</i> , <i>COL6A3</i> ,	Collagen chain trimerization	<i>COL6A6</i>
3	ADAMTS9, <i>COL4A2</i> , <i>COL6A3</i> ,	Degradation of the extracellular matrix	<i>COL6A6</i> , <i>LAMA5</i>
4	<i>COL4A2</i> , <i>COL6A3</i> ,	The activation of arylsulfatases	<i>COL6A6</i>
5	<i>COL4A2</i> , <i>COL6A3</i>	Collagen degradation	<i>COL6A6</i>
6	<i>COL4A2</i> , <i>COL6A3</i> ,	Assembly of collagen fibrils and other multimeric structures	<i>COL6A6</i> ,
7	<i>COL4A2</i>	Laminin interactions	<i>LAMA5</i>
8	<i>COL4A2</i> , <i>COL6A3</i> ,	Collagen biosynthesis and modifying enzymes	<i>COL6A6</i>
10	<i>COL4A2</i> , <i>COL6A3</i> , <i>SPTBN5</i>	NCAM signaling for neurite out-growth	<i>COL6A6</i>
11	<i>TMOD2</i>	Striated Muscle Contraction	
12	<i>COL4A2</i> , <i>COL6A3</i> ,	Collagen formation	<i>COL6A6</i> ,

S.No	Gene list from STU-66 family	Pathways	Gene list from STU-65 family
13		Interleukin-4 and Interleukin-13 signaling	<i>ITGAX, LAMA5</i>
14		Gamma carboxylation, hypusine formation and arylsulfatase activation	<i>ARSD</i>
15		Serotonin receptors	<i>HTR2B</i>
16	<i>COL4A2</i>	Anchoring fibril formation	
17	<i>COL4A2, COL6A3</i>	Integrin cell surface interactions	<i>COL6A6, ITGAX</i>
18	<i>RSC1A1</i>	Intestinal hexose absorption	
19	<i>ARSD</i>	Glycosphingolipid metabolism	
20	<i>COL4A2</i>	Crosslinking of collagen fibrils	
21	<i>CSGALNACT1</i>	Chondroitin sulfate biosynthesis	
22		Stimuli-sensing channels	<i>TRPV2, WNK1</i>
23	<i>RSC1A1</i>	Intestinal absorption	
24	<i>COL4A2, COL6A3</i>	Extracellular matrix organization	<i>COL6A6, ITGAX, LAMA5</i>
25		TRP channels	<i>TRPV2</i>

Note: *The candidate genes proposed are exclusive to affected segregating in heterozygous status

Supplementary tables:

Table A1: Compilation of the variants common in parent offspring pair of STU-66 family correlating to speech language disorders that includes 13 novel variants observed along with the bioinformatic predictions for pathogenicity

Table A2: Compilation of the variants common in sib pair of STU-65 family correlating to speech language disorders that includes 4 novel variants observed along with the bioinformatic predictions for pathogenicity

Table A3: Over-representation analysis results of variants identified in two ES families showing 25 most significant pathways related to stuttering using Reactome database

Table A4: Characterization of the stuttering phenotype and other associated findings among 24 individuals examined in the STU-65 family

Figure legends:

Figure 1: Five generation pedigree of stuttering family - STU-66

Figure 2: Six generational pedigree of a stuttering family - STU-65

Figure 3: Venn diagram showing the variants shared among the four individuals in the STU-66 family (15 combinations)

Figure 4: Flow chart depicting exome analysis of two multiplex stuttering families

Supplementary figures:

Figure A1: Bioinformatic-analysis pipeline applied to exome sequence data of a STU-66 family with stuttering

Figure A2: Venn diagram showing the variants shared between the affected sib pair in the STU-65 family

Figure A3: Bioinformatic-analysis pipeline applied to exome sequence data of a STU-65 family with stuttering

Figure A4: Venn diagram showing the genes shared between two stuttering multiplex families

Figure A5: Functional protein association network (STRING analysis) did not show any connection for seven proteins identified in common for the two multiplex stuttering families





