Molecular genetic examination of nonword repetition in a multigenerational family with a history of verbal trait disorders

Dongnhu T. Truong[1]; Andrew K. Adams[2]; Shelley D. Smith[3]; Lawrence D. Shriberg[4]; Jeffrey R. Gruen[1,2,5]

[1]Department of Pediatrics, Yale School of Medicine, New Haven, CT 06520; [2]Department of Genetics, Yale School of Medicine, New Haven, CT 06520; [3]Department of Pediatrics; University of Nebraska at Omaha, Omaha, NE 68182; [4]Waisman Center, University of Wisconsin-Madison, Madison, WI, 53705; [5]Investigative Medicine, Yale University School of Medicine, New Haven, CT 06520

Introduction

Verbal trait disorders encompass a wide range of conditions and are marked by deficits in five domains that impair a person's ability to communicate: speech, language, reading, spelling, and writing. Nonword repetition (NWR) is a robust endophenotype for verbal trait disorders. It has a strong genetic component and is sensitive to cognitive processes critical to verbal development, including auditory processing, phonological working memory, and motor planning and programming. In the present study, we examine a six-generation extended family of European ancestry with a history of verbal trait disorders. **We hypothesize** that genetic variants segregate with impaired NWR performance and that damaging mutations within these shared genomic regions contribute to the high prevalence of verbal trait disorders within the family.

Methods

Subjects:

62 individuals from a six-generation 90-member family of European ancestry with a history of verbal trait disorders. Written informed consent was approved by the University of Wisconsin-Madison Institutional Review Board.

Assessments:

Nonword Repetition Task (NRT; Dollaghan and Campbell 1998)
Syllable Repetition Task (SRT; Shriberg et al. 2009).

Phenotype Definition:

Individuals who scored less than 1 standard deviation below the mean of same age-sex speakers in a sample of 200 typically-speaking 3 to 80 year-old reference group participants (Potter 2012; Scheer-Cohen 2013) on either the NRT or SRT task were considered affected for a verbal trait disorder.

Genotyping:

551,839 single nucleotide polymorphism (SNP) markers were genotyped using the Illumina Infinium HumanCoreExome-24-v.1 at the Yale Center for Genome Analysis.

Statistical Analysis:

Genome-wide multipoint variance component linkage analyses were conducted to examine linkage between affected NWR performance and allele sharing (multipoint identity by descent) using SOLAR (Almasy and Blangero, 1998). Haplotypes were generated using MERLIN (Abecasis et al. 2002).

Targeted Deep sequencing

Capture oligos for the region of interest were synthesized using SeqCap EZ Choice XL (Roche). Targeted Deep sequencing for all 62 individuals was conducted by the Yale Center for Genomic Analysis. Raw sequencing data and variant calls were analyzed using GATK's best practices pipeline (DePristo et al., 2011). Analysis ready variants that passed GATK's QC were then annotated using ANNOVAR and VEP (Wang et al., 2010; McLaren et. al. 2016). Variants that were considered damaging by two or more variant annotation databases (SIFT, Poly Phen 2, MutationTaster, MutationAssessor, LRT, FATHMM, MetaSVM, MetaLR, and PROVEAN) were identified for further analysis.

Results

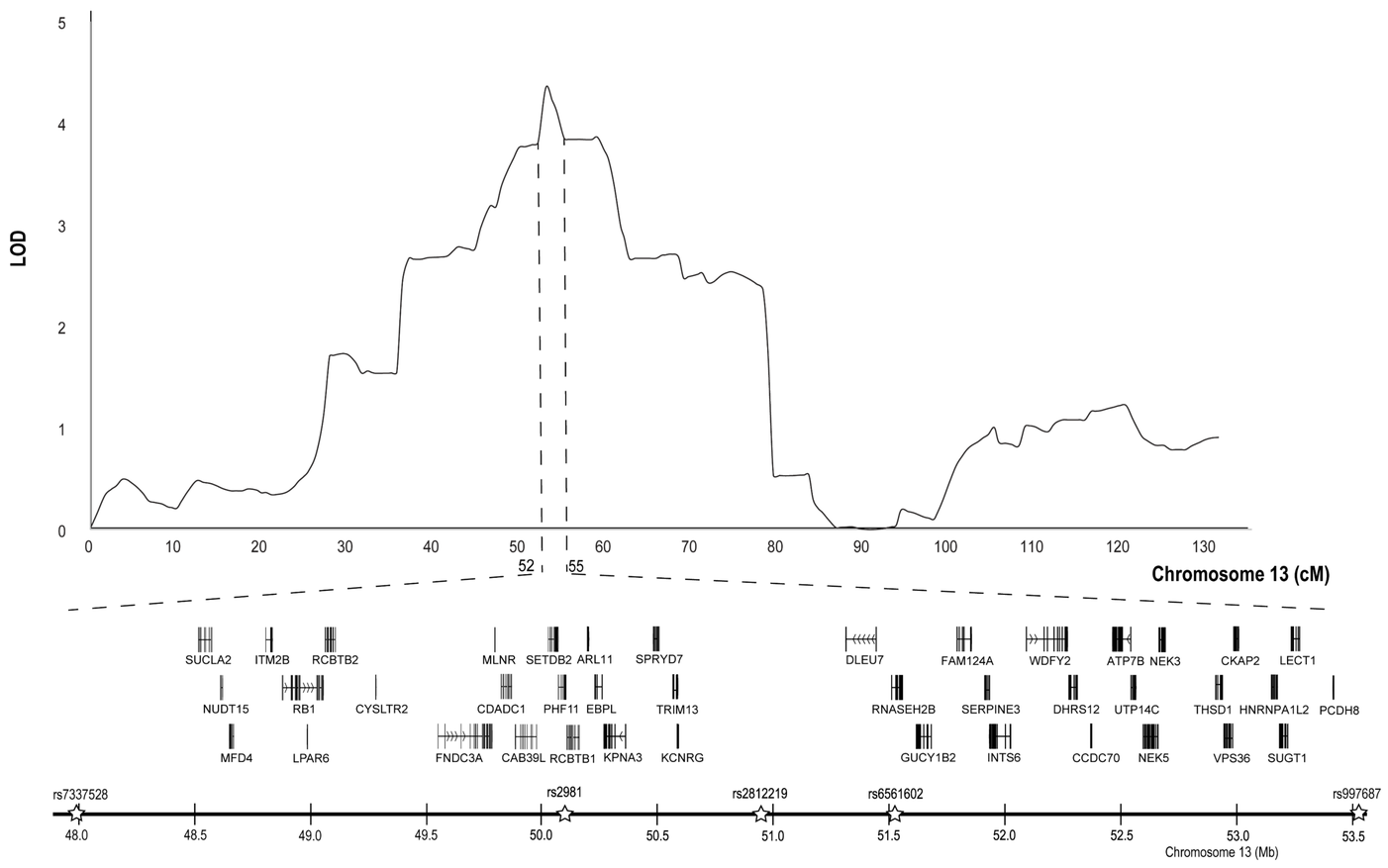

Figure 1: Multipoint linkage results conditioned on impaired NWR at chromosome 13. Genes and associated SNPs under the highest linkage peak of LOD = 4.35, between 52-55 cM spanning physical positions 48-53.5 Mb on reference genome assembly build GRCh37/hg19. The broader significant linkage signal (LOD > 3) lies between 45-62 cM spanning physical positions 43-70.5 Mb.

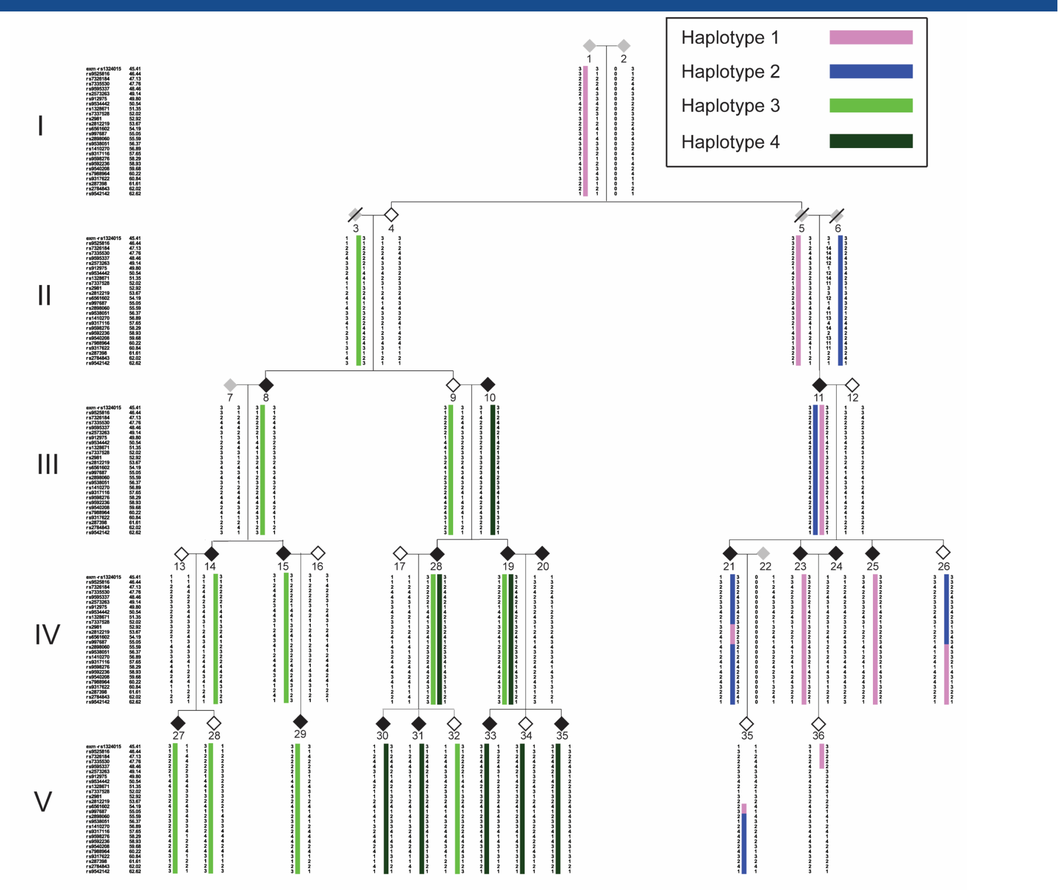

Figure 2: Haplotype assignments spanning genomic location 45-62 cM on chromosome 13. The pedigree depicted is truncated to reflect the three major haplotypes segregating with poor NWR in this family. A recombinatorial event observed in affected individual 21 outlines the centromeric and telomeric border of the linkage signal spanning 52-55 cM. Affected individuals are black diamonds, while unaffected individuals are grey diamonds.

Chr	Locus (cm)	Full Pedigree	(-) SubPed Hap1	(-) SubPed Hap3	(-) SubPed Hap4
13	50	3.7672	3.4394	1.8386	1.1119
13	51	3.787	3.4082	1.8913	1.1482
13	52	3.8089	3.386	1.9007	1.146
13	53	4.3465	3.3769	2.4413	1.6365
13	54	4.2213	3.3765	2.3585	1.5511
13	55	4.1128	3.3823	2.001	1.224
13	56	3.8518	3.3821	1.9479	1.1845
13	57	3.8396	3.3681	1.8933	1.1294
13	58	3.8391	3.318	1.8253	1.0457

Table 1 (LEFT): Multipoint linkage results (LOD scores) from the full pedigree and systematic exclusion of sub-pedigrees containing individuals with haplotype 1, 3, and 4, respectively. Results suggest that sub-pedigrees with haplotypes 3 and 4 largely contribute to the observed linkage signal on chromosome 13.

Chr	Position	Gene	Consequence	Minor Allele Frequency (NFE)	# Damaging calls (out of 10)	dbSNP	# Obs with Hap 3 (out of 10)	# Obs with Hap 4 (out of 8)	# Obs without Hap 3 or 4	# Obs Affected	# Obs Not affected
13	47345630	ESD	Missense variant; splice region variant	0.0158	5	rs15303	10	1	0	8	3
13	60566644	DIAPH3	Missense variant	0.053	5	rs36084898	10	0	0	7	3

Table 2 (ABOVE): Targeted sequencing results of observed damaging variants associated with Haplotype 3. Two predicted damaging variants in ESD (esterase D) and DIAPH3 (diaphanous-related formin 3) were observed in individuals with Haplotype 3.

Discussion

A genome-wide multipoint variance component linkage analysis of nonword repetition identified a region spanning chromosome 13q14-q21 with LOD = 4.35 between 52 and 55 cM, spanning approximately 5.5 Mb on chromosome 13.

- Region overlaps with SLI3, a locus implicated in reading disability in families with a history of specific language impairment

ESD (esterase D) encodes a S-formylglutathione hydrolase that metabolizes numerous substrates including O-acetylated sialic acids.

- O-acetylated sialic acids include gangliosides which are major membrane components of eukaryotic cells that play a role in cell signaling processes such as cell-cell interactions, cell motility, axonal sprouting, and neurite extension (Kohla et al., 2002).

DIAPH3 (diaphanous-related formin 3) encodes a member of the diaphanous subfamily of the formin family.

- Members of this family are involved in actin remodeling and regulate cell movement and adhesion (Higgs, 2002).
- Mutations in this gene are associated with autosomal dominant auditory neuropathy 1 (Kim et al., 2004).

Future Directions

- Identify variants that segregate with impaired NWR in noncoding regions of the genome using bioinformatic data to filter for regions of predicted tissue-specific functionality (CADD, GenoSkyline).

Acknowledgements

This study was supported by grant National Institutes of Health (NIH) R01 DC000496 (LDS), NIH HD03352 (Waisman Center), NIH R01 N5043530 (JRG), P50 HD027802 (JRG), and T32 HD07094 (DTT). We thank our colleagues in the Phonology Project, Waisman Center, for their assistance, and the staff at the Yale Center for Genome Analysis for genotyping and sequencing services We extend our sincere gratitude to a family member whose assistance and insights were central to this research, and to all participants for their contribution to this study.

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30: 97. doi: 0.1038/ng786

Almasy L, Blangero J (1998) Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. Am J Hum Genet 62: 1198-1211. doi: 10.1086/301844

Dollaghan C, Campbell TF (1998) Nonword Repetition and Child Language Impairment. J Speech Lang Hear Res 41: 1136-1146. doi: 10.1044/jslhr.4105.1136

Higgs HN. Formin proteins: a domain-based approach. *Trends Biochem Sci*2005; **30**: 342–353.

Kim, T. B., Isaacson, B., Sivakumaran, T. A., Starr, A., Keats, B. J. B., Lesperance, M. M. A gene responsible for autosomal dominant auditory neuropathy (AUNA1) maps to 13q14-21. J. Med. Genet. 41: 872-876, 2004.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F. (2016) The Ensembl Variant Effect Predictor. Genome Biology, 17:122

Potter NL, Hall, S., Karlsson, H. B., Fourakis, M., Lohmeier, H. L., McSweeny, J. L., Tilkens, C.M., Wilson, D.L., Shriberg, L. D. (2012)
Reference data for the Madison Speech Assessment Protocol (MSAP): A database of 150 participants 3-to-18 years of age with typical speech. . In: Project P (ed), Waisman Center, University of Wisconsin-Madison

Scheer-Cohen AR, Holt, A. S., Karlsson, H. B., Mabie, H. L., McSweeny, J. L., Tilkens, C. M., Shriberg, L. D. (2013) Reference Data for the Madison Speech Assessment Protocol (MSAP): A Database of Fifty 20-to-80 year old Participants with Typical Speech. In: Project P (ed), Waisman Center, University of Wisconsin-Madison

Shriberg LD, Lohmeier HL, Campbell TF, Dollaghan CA, Green JR, Moore CA (2009) A Nonword Repetition Task for Speakers with Misarticulations: The Syllable Repetition Task (SRT). J Speech Lang Hear Res 52: 1189-1212. doi: 10.1044/1092-4388(2009/08-0047)

Wang K, Li M, Hakonarson H. (2010) ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data Nucleic Acids Research, 38:e164