Identification of X-linked missense variants in *TAF1* in 5 unrelated families with autism spectrum disorder (ASD).

AUTHORS: Jonathan Lim[1,2], Kristina Calli[2,6], Ying Qiao[2,6], Lisa Hsieh[2,6], Sally Martell[2,6], Steven J.M. Jones[2, 3, 6], Stephen W. Scherer[4,6,7], M.E. Suzanne Lewis[2,6]

[1]Faculty of Medicine, University of British Columbia [2] Department of Medical Genetics, University of British Columbia [3]Michael Smith Genome Sciences Centre, Vancouver, BC, Canada [4]The Centre for Applied Genomics and McLaughlin Centre, Hospital for Sick Children, Toronto, Canada [5]BC Children's Hospital's Research Institute, Vancouver, Canada [6]iTARGET Autism (itargetautism.ca) [7]MSSNG (research.mss.ng)

Abstract

TAF1 (TATA-Box Binding Protein Associated Factor 1) is an X-linked gene encoding the largest component and core scaffold of the TFIID basal transcription factor complex. It plays an important role in neurodegenerative diseases and developmental delay. An increasing number of cases have been reported with variants in this gene grouped as a new neurodevelopmental syndrome (TAF1/MRXS33 intellectual disability syndrome). Common clinical features include hypotonia, facial dysmorphia, developmental delay, intellectual disability, and/or autism spectrum disorder (ASD). We hypothesized that *TAF1* plays a vital role in the contribution to this phenotype. Using parents-and-affected whole-genome sequencing methods, we identified 5 missense variants of *TAF1* in 5 boys among a cohort of 125 patients affected with ASD (p.Asp1496Ala, p.Pro215Leu, p.Leu619Phe, and p.Gly1391Arg). Out of the five probands, one showed a more severe phenotype (Complex-5) while the other four had either Simple-1 or Simple-2 according to our phenotype semi-quantitative evaluation criteria. Interestingly, the genomic location of the variant in the most severely affected patient was in the hot region of the gene. This area is enriched with a higher number of reported pathogenic/likely pathogenic variants. A detailed genotype-phenotype correlation analysis will be summarized among our patients and published cases in the future. Finally, other variants in multiple strong ASD candidate genes were found in each of our cases, inherited from both parents. Our data suggest that different types and locations of the variants in *TAF1* genes, as well as variants from different ASD candidate genes might contribute to the diversity of phenotypes in this syndrome.

Rationale

TAF1 (TATA-Box Binding Protein Associated Factor 1) is an X-linked gene that encodes the largest component and core scaffold of the TFIID basal transcription factor complex[1]. An increasing number of cases have been reported with the variants in this gene grouped as a new neurodevelopmental syndrome (TAF1/MRXS33 intellectual disability syndrome). The common clinical features appear early in life with hypotonia, facial dysmorphia, developmental delay, intellectual disability (ID) and/or autism spectrum disorder (ASD). The etiology of autism spectrum disorder continues to be extensively researched. There are many genes that have been shown to play a role in its development, and *TAF1* is a gene that may have a role, but the extent to which it plays a part is not clear yet.

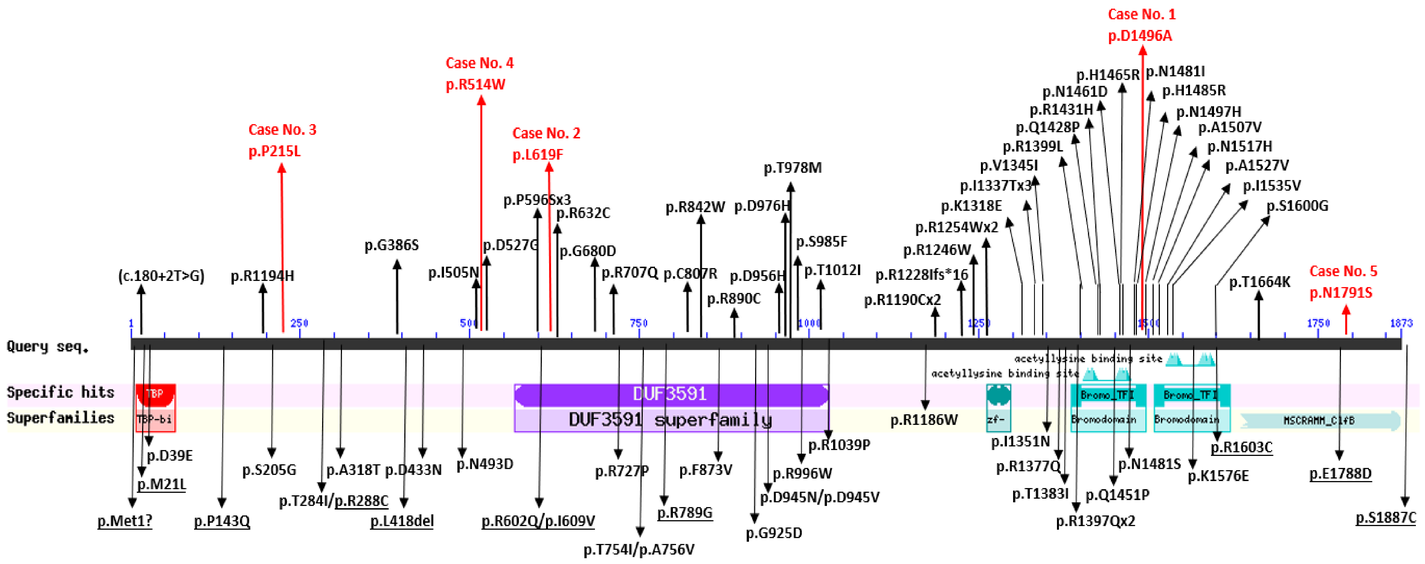
Methodology

Using parents-and-affected WGS method, we identified 5 rare missense variants of TAF1 in 5 individuals. *In silico* tools were used to predict the pathogenicity of the variants. To find other potentially deleterious variants that might contribute to the causes and/or susceptibility to ASD, we screened the WGS data in each family using the VarSeq annotation data, public databases and our internal pipelines. These discovered variants were shown to have high confidence criteria for being ASD genes as per the SFARI website. Finally, we compared the variants in our cases with variants reported in published papers and variants in Decipher database. It showed that the variant in Case No.1 (within Bromo-TFIID domain) is located within a hot spot for pathogenic or likely pathogenic variants, while the variants in Case No.3-5 are in regions surrounded by more VUS and/or likely benign variants (Figure 1). We also looked at the established literature that went into detail regarding *TAF1* variants and their phenotypes. Looking at the phenotypes of their cohorts, we compared them to our cohorts for similarities or differences (Table 1).

Results

Analysis showed that all five of the variants were hemizygous, maternally inherited from unaffected mothers. Of the five variants, the phenotypes of 4 of the five were either Simple-1 or Simple-2, while one was Complex-5. This was calculated according to our phenotype semi-quantitative evaluation criteria. We also found that in the more severely affected phenotype, the variant was in a hot spot for pathogenic or likely pathogenic variants. All 5 cases had a number of variant ASD genes discovered, the importance of each variant and its interaction with TAF1 is yet to be finalized. There were both similarities and differences between our cohort and those of the literature (Table 1), which will require further analysis.

Figure 1. Distribution of *TAF1* variants identified in our cases, published papers and Decipher database



The TAF1 protein diagram and domains were downloaded from NCBI. Variants of TAF1 identified from our five cases were marked red. The rest variants includes those reported in published papers with pathogeneicity description and variants in Decipher database (https://www.deciphergenomics.org/gene/TAF1/patient-overlap/snvs). The pathogenic and likely pathogenic variants were marked above the Query sequence bar; the variants with uncertain significance (VUS) and benign variants were below the bar; the benign or likely benign variants were below the bar and underlined. The two Bromo-TFIID domains have intervals as 1508-1620 and 1386-1497 amino acids, respectively.

Table 1: Phenotypic characteristics of our cohort in comparison with published literature

Phenotype characteristic (HPO ID)	Cheng 2019[2] (N=24) (%)	O'Rawe 2015[3] (N=14) (%)	iTarget cohort (N=5) (%)
Global developmental delay (0001263)	95.8	NR	0
Delayed speech and language development (0000750)	91.7	92.9	100
Delayed gross motor development (0002194)	66.7	100	0
Intellectual disability (0001249)	63.7	92.9	20
Seizures (0001250)	25.0	21.4	40
Gait disturbance (0001288)	29.2	64.3	0
GERD (0002020)	20.8	42.9	0
Microcephaly (0000252)	41.7	71.4	0
Self-injurious behaviour (0100716)	12.5	NR	60
Generalized hypotonia (0001290)	70.8	85.7	60
Feeding difficulties (0011968)	62.5	NR	40
Hypoplasia of corpus callosum (0002079)	22.7	71.4	0
IUGR (0001511)	29.2	35.7	0
High palate (0000218)	12.5	71.4	60

Future Directions

- Phenotypic clustering
- Correlation of phenotype with variant location
- Determining significance of *TAF1* variant alone, and effect of other ASD associated genes

Acknowledgements

We would like to extend our sincere thanks to the study participants/families involved in this study. iTARGET is supported by Genome British Columbia (Strategic Initiative B22ITG). The ASPIRE BC work was supported by funding from the Canadian Foundation for Innovation (CFI-LEF). MESL is a senior clinician scientist supported by the B.C. Children's Hospital Research Institute Investigator Grant Award Program (iGAP).