

Homozygous *TAFIC* variants are associated with a novel childhood-onset neurological phenotype

Running title: *TAFIC* Associated Neurological Phenotype

Oula Knuutinen, M.D.^{a,b*}; Angela Pyle, Ph.D.^{c*}; Maria Suo-Palosaari, M.D., Ph.D.^{b,d†}; Jennifer Duff, Ph.D.^{c†}; Tawfiq Froukh, Ph.D.^e; Anna-Elina Lehesjoki, M.D., Ph.D.^f; Salla M. Kangas, Ph.D.^{a,b,g}; James Cassidy, M.Sc.^c; Latifa Maraqa, M.D.^h; Riikka Keski-Filppula, M.D., Ph.D.^{a,b,i}; Hannaleena Kokkonen, Ph.D.^j; Johanna Uusimaa, M.D., Ph.D.^{a,b,k}; Rita Horvath, M.D., Ph.D.^{c§}; Päivi Vieira, M.D., Ph.D.^{a,b,k§}

^aPEDEGO Research Unit, University of Oulu, Finland

^bMedical Research Center, University of Oulu, Finland

^cWellcome Centre for Mitochondrial Research, Translational and Clinical Research Institute, Newcastle University, UK

^dDepartment of Diagnostic Radiology, Research Unit of Medical Imaging, Physics and Technology, Oulu University Hospital and University of Oulu, Finland

^eDepartment of Biotechnology and Genetic Engineering, Philadelphia University, Amman, Jordan

^fFolkhälsan Research Center and Medicum, University of Helsinki, Finland

^gBiocenter Oulu, University of Oulu, Finland

^hPKU & Genetic Consultation Clinic, Amman, Jordan

ⁱDepartment of Clinical Genetics, Oulu University Hospital, Finland

^jNorthern Finland Laboratory Centre NordLab, Oulu University Hospital, Finland

^kClinic for Children and Adolescents, Oulu University Hospital, Finland

*†§ Equal contribution.

Corresponding author:

Päivi Vieira, Clinic for Children and Adolescents, Oulu University Hospital, PO Box 23, 90029 Oulu, Finland. paivi.vieira@fimnet.fi.

Acknowledgements

This work was supported by the Arvo and Lea Ylppö Foundation, Stiftelsen Alma och K. A. Snellman Säätiö, and Oulun Lääketieteellinen Tutkimussäätiö.

Conflict of interest statement:

The authors declare no conflicts of interest.

Data availability statement:

The variants of patient 1 and 2 were submitted to RD-CONNECT (<https://rd-connect.eu/E503930>), and Leiden Open Variation Database (<https://databases.lovd.nl/shared/variants/0000667111#00020710>), respectively. Patients'

medical records are confidential. Other data supporting the findings are available from the authors upon reasonable request.

Abstract

TATA-box binding protein associated factor, RNA polymerase I subunit C (*TAFIC*) is a component of selectivity factor 1 belonging to RNA polymerase I (Pol I) transcription machinery. We report two unrelated patients with homozygous *TAFIC* missense variants and an early-onset neurological phenotype with severe global developmental delay. Clinical features included lack of speech and ambulation and epilepsy. MRI of the brain demonstrated widespread cerebral atrophy and frontal periventricular white matter hyperintensity. The phenotype resembled that of a previously described variant of *UBTF*, which encodes another transcription factor of Pol I.

TAFIC variants were located in two conserved amino acid positions and were predicted to be deleterious. In patient-derived fibroblasts, *TAFIC* mRNA and protein expression levels were substantially reduced compared to healthy controls. We propose that the variants impairing *TAFIC* expression are likely pathogenic and relate to a novel neurological disease. This study expands the disease spectrum related to Pol I transcription machinery, associating the *TAFIC* missense variants with a severe neurological phenotype for the first time.

Keywords: [brain atrophy](#), [infantile spasms](#), [PolI transcription initiation complex proteins](#), [TATA-binding protein associated factors](#)

Introduction

TATA-box binding protein associated factor, RNA polymerase I subunit C (*TAF1C*, MIM 604905) belongs to the RNA polymerase I (Pol I) transcription machinery producing ribosomal RNA (rRNA).¹ In mammals, two transcription factors are required for Pol I to initiate transcription, selectivity factor 1 (SL1) and upstream binding factor (UBF) encoded by *UBTF* (MIM 600673). SL1 is a complex formed by TATA-binding protein (TBP, MIM 600075) and three Pol I-specific TBP-associated factors (TAFs).² Recently, a variant of *UBTF* leading to gain-of-function of UBF has been reported to cause altered rRNA metabolism and a paediatric neurodevelopmental disorder.³⁻⁶ The clinical picture is characterised by developmental regression and brain atrophy leading to microcephaly.

Here we report two homozygous variants of *TAF1C* associated with neurological disease. A clinical description of the patients is provided together with molecular studies.

Materials and methods

Study subjects

The index patient was examined at the Clinic for Children and Adolescents of Oulu University Hospital, Finland (patient 1). Another patient examined in Jordan (patient 2) was identified via Genematcher.⁷ Informed consent was obtained from all subjects or their guardians. This study was carried out in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Northern Ostrobothnia Hospital District and Newcastle and North Tyneside 1 National Research Ethics Committee.

Molecular genetic studies

For patient 1, after excluding genes underlying PEHO syndrome (progressive encephalopathy with oedema, hypsarrhythmia, optic atrophy, MIM 260565) and PEHO-like phenotypes (*ZNHIT3*, *UBA5*, *ADAM22* and *SEPSECS*)⁸⁻¹¹ by Sanger sequencing (University of Helsinki, Finland), and copy number variants by chromosomal microarray (Supporting information), exome sequencing was carried out in DNA extracted from the patient and parents (Newcastle University, UK). Exome capture was obtained using the Nextera Rapid Exome Capture (Illumina, USA), sequenced with paired-end reads on an Illumina NextSeq500 and aligned to the human reference genome (UCSC hg38). Homozygous or compound heterozygous exonic or splice-site variants were prioritised with a minor allele frequency less than or equal to 0.01 in external databases including gnomAD. Variant pathogenicity was assessed using *in silico* prediction tools including SIFT, Polyphen2, and CADD. Variants were verified in the patient and the parents using Sanger sequencing. For patient 2, the exome sequencing was performed in an accredited clinical genetic laboratory (Macrogen Inc., Seoul, South Korea, Supporting information). The reference sequence NM_005679.3 for *TAF1C* was used in all analyses.

Western blot and quantitative real-time PCR

Expression of *TAF1C* was analysed from primary skin fibroblasts of patient 1 and three healthy controls by Western blot (Supporting information) and quantitative real-time PCR. RNA from fibroblasts was extracted using the RNeasy mini kit (Qiagen, Manchester, UK) and cDNA was generated using High-Capacity cDNA Reverse Transcription Kit (Life Technologies Ltd, Paisley, UK). Quantitative real-time PCR (qPCR) was performed using iTaq™ Universal SYBR® Green Supermix on a CFX96 Touch™ PCR system (BioRad,

Hertfordshire, UK). *TAF1C* gene expression levels were measured relative to *GAPDH* and β -*Actin*.

Results

Patient 1

Patient 1 was the fourth child of unaffected, consanguineous Finnish parents (Fig. 1A). She was born at week 38 by induced labour due to gestational diabetes. The mother's glucose tolerance test was repeatedly normal after pregnancy. The Apgar scores were 9/9, head circumference at 50th percentile, and weight at 80th percentile (3650 grams). She needed phototherapy for transient hyperbilirubinemia and an enlarged liver was noted on abdominal ultrasound, which later normalised (Table 1).

At the age of three months, she was referred to the paediatric neurologist because of poor eye contact and opisthotonus. She did not reach any motor milestones but developed a social smile and transient eye contact at the age of six months after prescription of corrective lenses. At seven months, epileptic spasms and hypsarrhythmia developed. Seizures continued despite multiple antiepileptic drugs. Valproate, levetiracetam, vigabatrin, topiramate, and nitrazepam improved seizure control, but effects were not sustained. Oxcarbazepine was also tried with minimal effect. At 18 months, she was diagnosed with central precocious puberty and leuprolide treatment was started. A gastrostomy tube was inserted at 3.5 years of age. The modified ketogenic diet was started thereafter with a sustained improvement of seizures.

Patient 2

Patient 2 was born to first-cousin parents of Jordanian and Syrian descent (Fig. 1B). Her birth weight was at 50th percentile (3500 grams). She had global developmental delay, spastic

paraplegia, hypotonia, hyporeflexia, dysmorphic features, and speech limited to few words. She had a history of head nodding and ataxia.

Radiological findings

For both patients, brain magnetic resonance imaging (MRI) at the age of six months was normal (Fig. 2, Panel 1 and 4). Follow-up MRI of patient 1 at the age of 17 and 36 months (Fig. 2, Panel 2 and 3) revealed diffuse atrophy and white matter abnormalities. No follow-up MRI of patient 2 was available.

Molecular findings

Exome sequencing identified homozygous *TAF1C* missense variants c.1165C>T p.Arg389Cys (rs140327311) in patient 1 and c.1213C>T p.Arg405Cys (rs759412906) in patient 2 (segregation shown in Fig. 1A–B, variant locations in 1G). Both variants were rare (frequency<0.002) in gnomAD and no homozygous variants were present. Only the variant of patient 1 was present in the Finnish SISu database (frequency of 0.004). The variants were located in a highly conserved area (Fig. 1C) and predicted to be deleterious by *in silico* prediction tools (Table 1). qPCR and western blot analyses revealed *TAF1C* mRNA and protein expression to be substantially reduced in fibroblasts of patient 1 compared to controls (Fig. 1D–F).

Discussion

We describe two patients with different homozygous missense *TAF1C* variants (Table 1) associated with severe early-onset neurological phenotype. The pathogenicity is supported by markedly reduced *TAF1C* mRNA and protein expression in fibroblasts of patient 1. Both

variants were rare in gnomAD, predicted to be deleterious, and affect highly conserved amino acid residues. This is the first known association between *TAFIC* and a neurological phenotype.

Previously, *TAFIC* has been linked only with type 2 diabetes¹² and gastrointestinal cancers.¹³ As Pol I transcription machinery produces ribosomal RNA driving cellular metabolism, it has been hypothesised that ribosomal stress due to altered Pol I activity could cause microcephaly and leukodystrophy through a tumour protein p53 (*TP53*) mediated mechanism (reviewed by Hetman and Slomnicki).¹⁴ Defects of ribosomal biogenesis can lead to the accumulation and activation of p53 followed by cell cycle arrest and induction of apoptosis. This mechanism might be a contributing factor to microcephaly and white matter abnormalities seen in patient 1. The *TAFIC* variants described in the current work were located in a predicted WD40-repeat domain (Fig. 1G), which have been shown to affect neuronal morphology and corpus callosum genesis.¹⁵ Interestingly, thinning of the corpus callosum was observed in patient 1.

Patient 1 was initially suspected of PEHO syndrome which belongs to the Finnish disease heritage. Four of five diagnostic criteria were present (infantile-onset hypotonia, hypsarrhythmia, profound intellectual disability, and progressive brain atrophy), but optic atrophy was not observed. Known variants related to PEHO syndrome⁸ and PEHO-like phenotypes⁹⁻¹¹ were sequenced but no pathogenic variants were found.

A variant of *UBTF* encoding another transcription factor of Pol I (UBF) has recently been associated with neurodegenerative disease in children.³⁻⁶ The main clinical features were highly similar to the phenotype of the patients in this study (Table 1). After being initially normal, brain MRI of patient 1 showed white matter hyperintensities in T2-weighted images,

thinning of the corpus callosum, and diffuse brain atrophy (Fig. 2) which have also been reported with the *UBTF* variant. In the *UBTF* related disorder, gain-of-function of UBF leads to increased expression of rRNA.^{3,4} In contrast, our study demonstrates a severe reduction of *TAFIC* expression (Fig. 1F) suggesting a loss-of-function mechanism of *TAFIC*, leading to impairment of Pol I transcription. Research on rRNA components in *TAFIC*-related disease may expand the knowledge of Pol I disorder pathomechanisms.

Based on these findings, we suggest that the homozygous *TAFIC* variants reported here are likely pathogenic and disease-causing in a recessive manner and thus expand the phenotype of neurological disorders associated with RNA polymerase I machinery. Further studies are needed to better describe the pathogenesis and novel neurological phenotype related to *TAFIC*.

References

1. Goodfellow SJ, Zomerdijk JCBM. Basic Mechanisms in RNA Polymerase I Transcription of the Ribosomal RNA Genes. In: Sub-Cellular Biochemistry. 2013;61:211-236.
2. Russell J, Zomerdijk JCBM. The RNA polymerase I transcription machinery. Biochem Soc Symp. 2006;(73):203-216.
3. Edvardson S, Nicolae CM, Agrawal PB, et al. Heterozygous De Novo UBTF Gain-of-Function Variant Is Associated with Neurodegeneration in Childhood. Am J Hum Genet. 2017;101(2):267-273.
4. Toro C, Hori RT, Malicdan C V, et al. A recurrent de novo missense mutation in UBTF causes developmental neuroregression. Hum Mol Genet. 2018;27(4):691-705.
5. Sedláčková L, Laššuthová P, Štěrbová K, et al. UBTF Mutation Causes Complex Phenotype of Neurodegeneration and Severe Epilepsy in Childhood. Neuropediatrics. 2019;50(1):57-60.
6. Bastos F, Quinodoz M, Addor MC, et al. Childhood neurodegeneration associated with a specific UBTF variant: a new case report and review of the literature. BMC Neurol. 2020;20(1):17.
7. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Hum Mutat. 2015;36(10):928-930.
8. Anttonen A-K, Laari A, Kousi M, et al. ZNHIT3 is defective in PEHO syndrome, a severe encephalopathy with cerebellar granule neuron loss. Brain. 2017;140(5):1267-

- 1279.
9. Muona M, Ishimura R, Laari A, et al. Biallelic Variants in UBA5 Link Dysfunctional UFM1 Ubiquitin-like Modifier Pathway to Severe Infantile-Onset Encephalopathy. *Am J Hum Genet.* 2016;99(3):683-694.
 10. Muona M, Fukata Y, Anttonen A-K, et al. Dysfunctional ADAM22 implicated in progressive encephalopathy with cortical atrophy and epilepsy. *Neurol Genet.* 2016;2(1):e46.
 11. Anttonen A-K, Hilander T, Linnankivi T, et al. Selenoprotein biosynthesis defect causes progressive encephalopathy with elevated lactate. *Neurology.* 2015;85(4):306-315.
 12. Abdulwahab RA, Allaith AAA, Shinwari Z, Alaiya A, Giha HA. Association of TATA box-binding protein-associated factor RNA polymerase I subunit C (TAF1C) with T2DM. *Gene.* 2019;706:43-51.
 13. Oh HR, An CH, Yoo NJ, Lee SH. Frameshift mutations of TAF1C gene, a core component for transcription by RNA polymerase I, and its regional heterogeneity in gastric and colorectal cancers. *Pathology.* 2015;47(2):101-104.
 14. Hetman M, Slomnicki LP. Ribosomal biogenesis as an emerging target of neurodevelopmental pathologies. *J Neurochem.* 2018;148(3): 325-347.
 15. Kannan M, Bayam E, Wagner C, et al. WD40-repeat 47, a microtubule-associated protein, is essential for brain development and autophagy. *Proc Natl Acad Sci U S A.* 2017;114(44):E9308-E9317.

Fig. 1. Homozygous *TAF1C* variants cause decreased expression of *TAF1C* mRNA and protein in patient-derived skin fibroblasts.

A. Family pedigree of patient 1. **B.** Family pedigree of patient 2. **C.** Conservation of *TAF1C* showing the amino acid positions of the reported variants. **D.** Western blot demonstrates decreased expression of *TAF1C* in fibroblast sample of patient 1 compared to healthy controls. **E.** *TAF1C* protein expression in relation to *GAPDH* is reduced in patient 1 compared to controls. **F.** Quantitative PCR of *TAF1C* mRNA in relation to *GAPDH* and β -*Actin*. **G.** Schematic presentation of *TAF1C* transcript NM_005679.3 (above) and *TAF1C* protein with predicted structural domains (below) in relation to the location of reported variants.

Fig. 2. Brain MRI findings in patients with *TAF1C* variants.

Panels 1–3: patient 1. Panel 4: patient 2.

Panel 1: T1-weighted (T1W) midsagittal (A), T1W coronal (B), T2-weighted (T2W) axial (C), and T2 Fluid Attenuated Inversion Recovery (FLAIR) axial (D) images show normal brain structure and myelination.

Panel 2: T1W midsagittal (E) and T2W coronal (F) images demonstrate mild cerebral, cerebellar, and brain stem atrophy. The quadrigeminal cistern is slightly dilated (E, arrow) and the corpus callosum is thinner than normal (E, arrowhead). T2W axial image (G, arrows) shows atrophy of the cerebellar peduncles. Frontal periventricular white matter hyperintensities are demonstrated on T2 FLAIR coronal image (H, arrows).

Panel 3: T1W midsagittal (I), T2W coronal (J), and T2W axial (K) images show progressive atrophy of the pontocerebellar structures (I–K), but the cerebral atrophy has not progressed. Dilatation of the quadrigeminal cistern (I, arrow), atrophy of the corpus callosum (I, arrowhead), and periventricular white matter hyperintensities (L, arrows) remain unchanged.

Panel 4: T1W midsagittal (M) and T2W coronal (N) images show normal corpus callosum and no atrophy of the brain. The cerebellar peduncles are normal on T2W axial image (O, arrows). T2W axial image (P) demonstrates normal myelination and no white matter signal abnormalities.

Table 1. Clinical characteristics of patients with *TAFIC* variants compared with the *UBTF* related disorder.

Table 1. Clinical characteristics of patients with *TAF1C* variants compared with the *UBTF* related disorder.

	Patient 1	Patient 2	<i>UBTF</i> related disorder (13 patients) ³⁻⁶
Current age	7 years	8 years	n/a
Age at onset	Neonatal	4 months	2–4 years
Gender	Female	Female	8/13 female
Ethnicity	Caucasian	Mixed Syrian-Algerian	Various
Motor development	Nonambulatory	Delayed	9/13 nonambulatory
Speech development	Nonverbal	Limited to few words	13/13 nonverbal or severe language impairment
Cognitive development	Intellectual disability, profound	Intellectual disability, moderate	13/13 intellectual disability
Muscle tone	Spastic, axial hypotonia	Spastic, transient hypotonia	Variable spasticity and hypotonia
Hearing	Normal	Normal	n/a
Ophthalmic findings	Strabismus, hyperopia	Strabismus, nystagmus	n/a
Head circumference	-2.3 SD	-1.1 SD	7/12 microcephalic
Endocrine findings	Precocious puberty	None	None
Cardiac findings	First-degree AV block	None	None
Liver findings	Transient neonatal hepatomegaly, normal ultrasonography at ages 1, 4, and 6 yrs. Transient ALT elevation up to 78 U/l [†] after implementation of the ketogenic diet.	Normal ALT levels at 2 yrs.	n/a
Dentition	Delayed	Normal	Normal
Dysmorphic features	None	Hypertelorism, depressed nasal bridge, frontal bossing	None
Seizures	Focal, generalised, spasms	No	5/13 focal or generalised
EEG findings	Intermittent hypsarrhythmia, slow background	None	7/13 abnormal
Brain MRI findings	Normal at 6 mos. Cerebral and cerebellar atrophy, T2 white matter hyperintensity at 17 and 36 mos.	Normal at 6 mos	13/13 cerebral and cerebellar atrophy, white matter abnormalities
Genetic variant	<i>TAF1C</i> NM_005679.3: c.1165C>T p.Arg389Cys, homozygous	<i>TAF1C</i> NM_005679.3: c.1213C>T p.Arg405Cys, homozygous	<i>UBTF</i> NM_014233.3: c.628G>A p.Glu210Lys, heterozygous
SIFT score	0.02 (deleterious)	0.00 (deleterious)	0.51 (tolerant)
Polyphen-2 score	0.998 (probably damaging)	0.998 (probably damaging)	0.999 (probably damaging)
CADD score	32.0	27.8	32.0

[†] ALT reference range <40 U/l. Abbreviations: ALT, alanine aminotransferase; AV, atrioventricular; EEG, electroencephalography; mos, months; MRI, magnetic resonance imaging; n/a, not available; SD, standard deviation; yrs, years.

Fig. 1

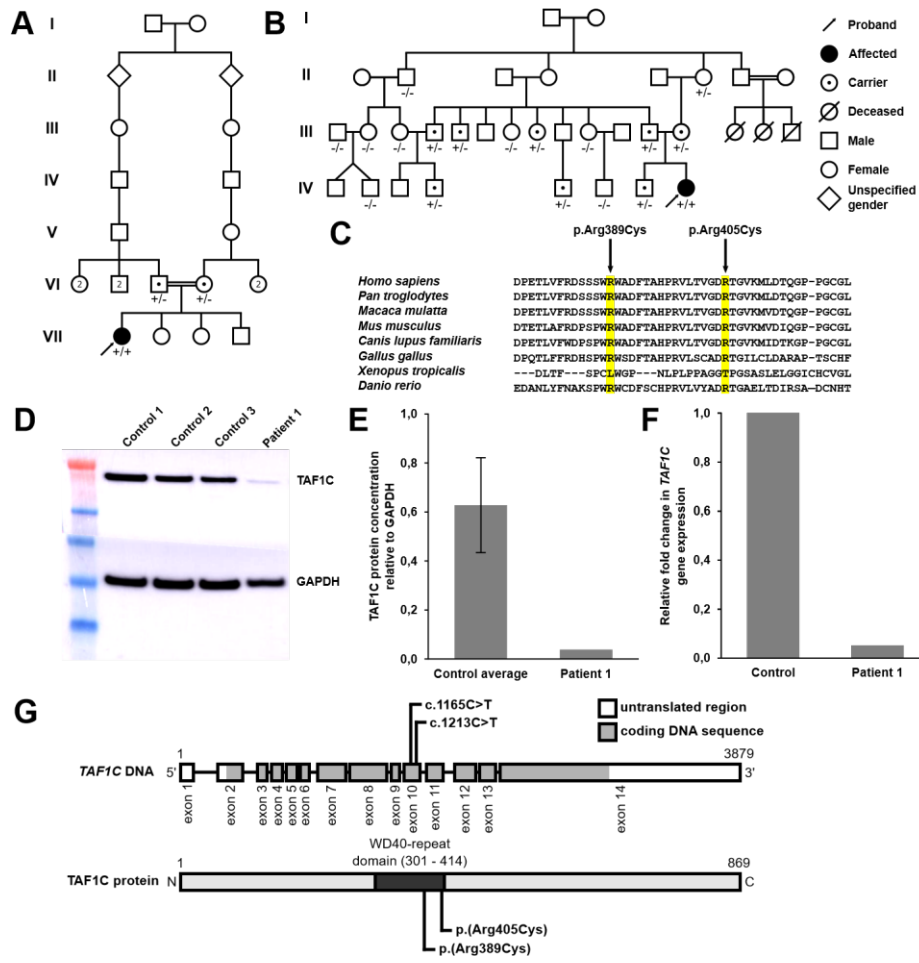
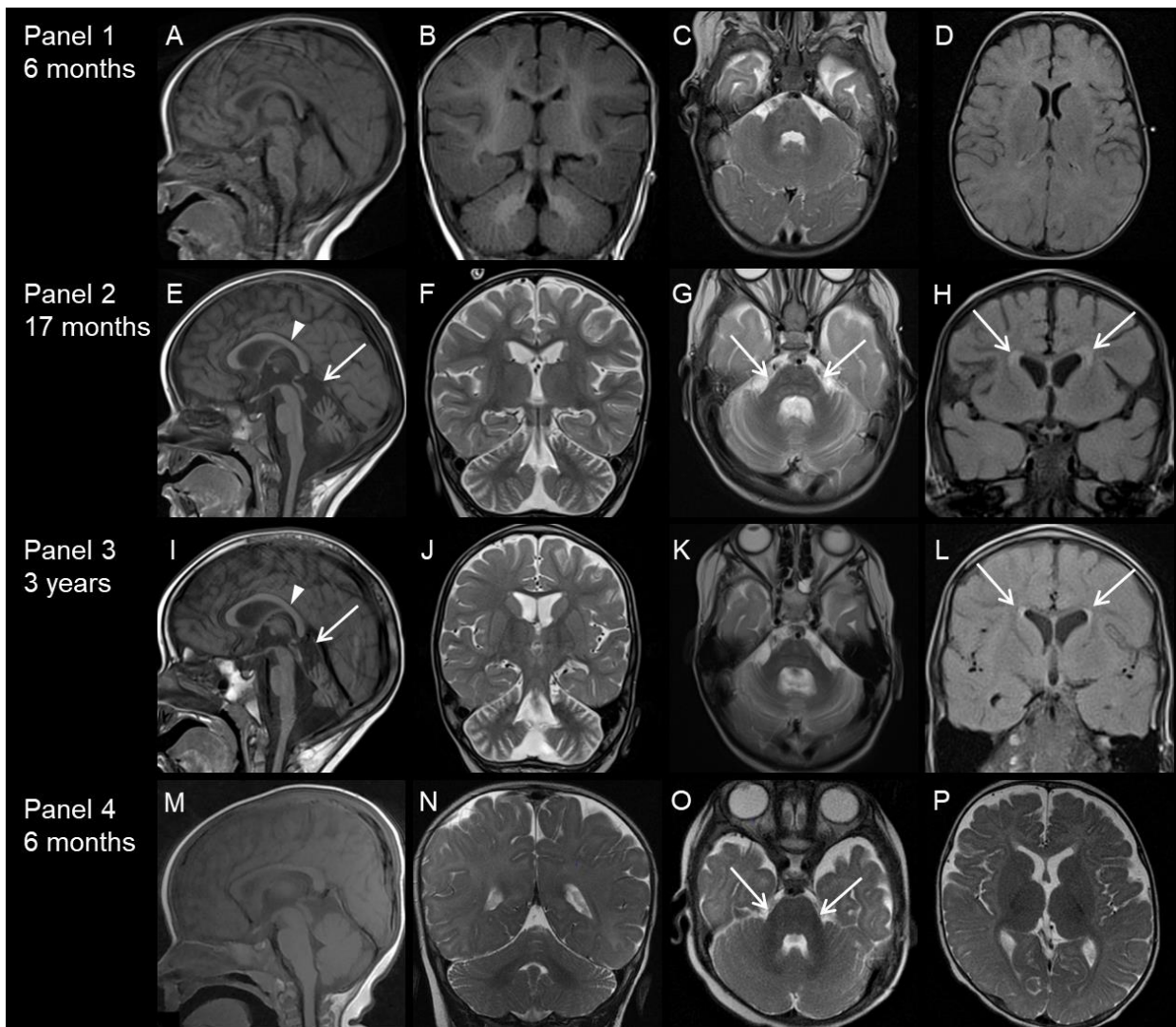


Fig. 2



Supporting information

Homozygous TAF1C variants are associated with a novel childhood-onset neurological phenotype

Oula Knuutinen, M.D.; Angela Pyle, Ph.D.; Maria Suo-Palosaari, M.D., Ph.D; Jennifer Duff, Ph.D.; Tawfiq Froukh, Ph.D.; Anna-Elina Lehesjoki, M.D., Ph.D.; Salla M. Kangas, Ph.D.; James Cassidy, M.Sc.; Latifa Maraqa, M.D.; Riikka Keski-Filppula, M.D., Ph.D; Hannaleena Kokkonen, Ph.D.; Johanna Uusimaa, M.D., Ph.D.; Rita Horvath, M.D., Ph.D.; Päivi Vieira, M.D., Ph.D.

Supplemental methods

Exome sequencing

For the family of patient 2, total genomic DNA was extracted with FlexiGene DNA Kit (Qiagen, Germany). For exome capture, SureSelect^{XT} V.6 Library Prep Kit was used (Agilent Technologies, USA). The Illumina NOVASeq6000 platform (Illumina, USA) was employed for the exome sequencing. The sequencing reads (150 bp pair end) were mapped to the reference genome (UCSC hg19) using the Burrows-Wheeler Aligner software.¹ Polymerase chain reaction duplicates were removed using SAMBLASTER.² Single-nucleotide variants and small insertions/deletions (indels) were called using FreeBayes and annotated using SnpEff version 3.3³ (Ensembl GRCh37). Sequencing was conducted by Macrogen (Seoul, South Korea) and the pipeline megSAP at Philadelphia University, Jordan, was used.

Western blot

Total protein lysates prepared from primary skin fibroblasts of patient 1 and three healthy controls were separated on NuPAGETM NovexTM 4-12% Bis-Tris Protein Gels. Following electrophoresis, the separated proteins were transferred onto PVDF membrane using the iBlot[®]2 Dry Blotting System (Thermo Fisher Scientific, UK) according to the manufacturer's

instructions. The membrane was then probed for the following proteins: TAF1C (Abcam, ab210750) and GAPDH (Santa Cruz, sc-25778).

Microarray analysis

For patient 1, total genomic DNA was extracted from 500 µL of peripheral blood using the QiaSymphony (Qiagen, USA) according to the manufacturer's instructions. Chromosomal microarray was performed using 50mer oligochip (HumanCytoSNP-12 (v2.1), Illumina, USA) allowing the effective resolution of >30 kB in cytogenetically relevant regions and >200 kB in other areas of the genome. Copy-number changes and regions of single nucleotide polymorphisms homozygosity were analyzed with GenomeStudio v.2011.1 and KaryoStudio 1.3 programs (Illumina, USA) and compared with known copy-number variations and gene content listed in the database of genomic variants (DGV), dbVar, UCSC genome browser, DECIPHER, ISCA and OMIM databases, and peer-reviewed literature searches in the PubMed database. Genomic positions were defined according to the human reference genome GRCh37/hg19.

Supplemental results

Microarray analysis

In patient 1, chromosomal microarray analysis (Fig. S1) revealed an interstitial 752.9 kB duplication of 6q27 (arr(GRCh37) 6q27(168570657_169323526)x3). The duplication region involves only two protein-coding genes, *SMOC2* (MIM 607223) and *DACT2* (MIM 608966) of which *SMOC2* is associated with tooth abnormalities (Dentin dysplasia, type I, with microdontia and misshapen teeth, autosomal recessive, MIM 125400). *DACT2* gene is involved in the regulation of intracellular signalling and is not known to cause any disease. DGV and dbVar showed three overlapping duplications of which one was interpreted as a

variant of unknown significant (VUS) and two as benign. Few additional variants overlapping either *SMOC2* or *DACT2* gene were also classified as benign. In patient databases, two corresponding 6q27-duplications have been described of which one was interpreted as benign (patient nssv584799, ISCA) and other as VUS (patients 265407, Decipher). Analysis of parents showed that the 6q27 duplication was inherited from the healthy father and this duplication was classified as likely benign (not associated with severe symptoms of this patient).

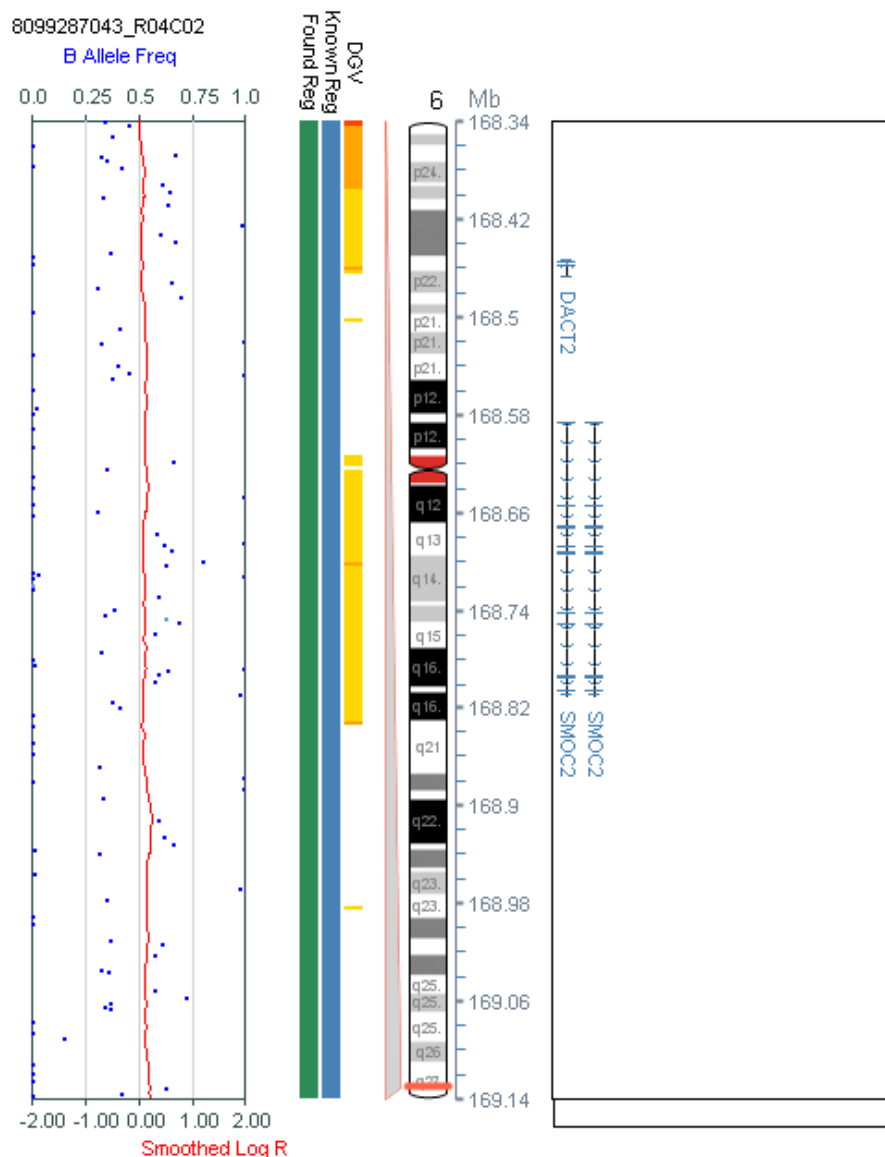


Fig. S1. Chromosomal microarray analysis of patient 1.

Supporting references

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
2. Faust GG, Hall IM. SAMBLASTER: Fast duplicate marking and structural variant read extraction. *Bioinformatics*. 2014;30(17):2503–5.
3. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* (Austin). 2012;6(2):80–92.