

1 **Mutations in *ZBTB20* in individuals with persistent stuttering**

2
3 Carlos E Frigerio Domingues¹⁺, Muhammad Hashim Raza^{1+^}, Tae-Un Han^{1#}, Terra Barnes², Philip Shaw³,
4 Gustavo Sudre³, Sheikh Riazuddin⁴, Robert J. Morell⁵, Dennis Drayna^{1*}

5
6 1. Laboratory of Communication Disorders, National Institute on Deafness and Other Communication
7 Disorders, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

8 2. Department of Neuroscience, School of Medicine, Washington University in St. Louis, MO

9 3. Neurobehavioral Clinical Research Section, National Human Genome Research Institute, National
10 Institutes of Health, Bldg. 31, Bethesda MD 20892

11 4. Allama Iqbal Medical College, University of Health Sciences, Lahore, Pakistan

12 5. Genomics and Computational Biology Core, National Institute on Deafness and Other Communication
13 Disorders, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

14 + These authors contributed equally to this research

15 # Current address: Molecular Neurogenetics Section, National Human Genome Research Institute,
16 National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

17 % Current address: Utah Center for Genetic Discovery, Eccles Institute of Human Genetics, University of
18 Utah, Salt Lake City, UT 84112

19 * To whom correspondence should be addressed at: NIDCD/National Institutes of Health
20 Porter Neuroscience Research Center
21 35A Convent Drive, Bethesda, MD 20892
22 Tel. 301-221-4392
23 drayna@nidcd.nih.gov
24 ORCID 0000-0002-1730-4532
25

26

27

28

29 **Key words:** Stuttering, Genetics, *ZBTB20*, Astrocytes

30

31

32

33

34

35

36

37

38

39

1 **Abstract**

2

3 Background: Previous studies identified a strong linkage signal for non-syndromic persistent
4 developmental stuttering on chromosome 3q13.2-3q13.33 in a large consanguineous family. To identify
5 the causative genetic variant at this locus, including we performed further analysis, including whole
6 exome, whole genome, and targeting Sanger sequencing.

7 Results: We identified a homozygous rare c.2155G>A variant in *ZBTB20* in individuals who stutter. This
8 mutation encodes Isoleucine in place of a highly conserved Valine at amino acid position 719 in ZBTB20
9 that co-segregates (LOD = 4.23) with stuttering under a model of recessive inheritance with reduced
10 penetrance in this family. Coding variants in this gene were significantly more frequent in a multiethnic
11 cohort of unrelated individuals who stutter than in individuals in large population databases comprised of
12 our normal control subjects and gnomAD database subjects matched for ethnicity. *ZBTB20* encodes a
13 zinc-finger transcription factor, and luciferase reporter constructs using genes known to be regulated by
14 ZBTB20 showed that the Ile719 mutant form of the protein displays altered transcriptional regulation *in*
15 *vitro*. Although homozygosity for mutations in *ZBTB20* has not previously been observed, dominant
16 mutations in *ZBTB20* have been reported in Primrose syndrome, a rare Mendelian dominant disorder
17 characterized by tall stature, developmental delay, dysgenesis of the corpus callosum, but generally not
18 speech disorders. Clinical re-examination of the affected family members ten years after their original
19 ascertainment revealed only some suggestive signs associated with Primrose syndrome.

20 Conclusions: Our findings support variants in *ZBTB20* as a cause of stuttering in a large family, and
21 rarely in the wider population. Previous studies in mice have demonstrated that *Zbtb20* is required for the
22 development of astrocytes, a brain cell type previously implicated in an animal model of stuttering. Our
23 findings support our hypothesis that astrocyte pathology is involved in this disorder. Our findings also
24 broaden the medical genetic view of disorders of *ZBTB20*, and demonstrate that individuals carrying

1 a homozygous mutation in this gene can be viable, and that they can primarily display persistent
2 developmental stuttering.

3 **Introduction**

4 Stuttering is a common neurodevelopmental speech disorder that affects the fluency of speech,
5 and is characterized by repetitions and prolongations of speech sounds, and by silent interruptions in the
6 flow of speech known as blocks [1,2]. Twin and adoption studies indicate a high heritability for this
7 disorder [3], however Mendelian segregation does not typically occur [4] and stuttering is considered to
8 be a complex genetic disorder. Linkage studies in consanguineous and polygamous families have
9 identified a number of loci that display strong linkage to this disorder [5,6,7,8,9], and causative genes for
10 stuttering have been identified at a number of these loci [10,11]. The products of the genes identified thus
11 far all physically interact and participate in functions important to intracellular trafficking. Animal studies
12 aimed at elucidating the cellular and anatomic neuropathology caused by these mutations have recently
13 revealed deficits in vocalization and abnormalities in astrocytes, particularly in the corpus callosum
14 [12,13].

15 We have previously reported a large consanguineous family in which many members are affected
16 by persistent neurodevelopmental stuttering. Linkage analysis identified significant linkage on
17 chromosome 3 under an autosomal recessive model with reduced penetrance [8]. Here we have performed
18 additional linkage analyses as well as whole exome, whole genome, and Sanger sequencing, which
19 revealed a unique rare homozygous coding variant in *ZBTB20* (MIM: 606025) that co-segregates with
20 stuttering in this family. Although no individuals with homozygous mutations in *ZBTB20* have previously
21 been observed, heterozygous mutations have been associated with Primrose syndrome, a rare multisystem
22 developmental disorder [14, 15, 16, 17, 18, 19, 20, 21, 22, 23].

23 *ZBTB20*, also known as *DPZF* [24], *HOF* [25], or *ZNF288* is a member of the POK (POZ and
24 Krüppel) family of transcription factors. We sought to determine if this variant exerted a functional effect
25 on transcription using a luciferase reporter construct in transfected cells. We also performed additional

1 clinical evaluations of the affected subjects in this family, and constructed and tested ultrasonic
2 vocalizations in a knock-in mouse line that carries this mutation.

3 **Results**

4 Stuttering in Family PKST77

5 Consanguineous family PKST77 has been previous described [8]. Speech abnormalities in all five
6 affected subjects were consistent with typical, non-syndromic persistent developmental stuttering. Based
7 on our subsequent genetic findings, family members were re-evaluated 10 years after their initial
8 enrollment. Clinical examinations of four available affected family members were performed at the
9 Allama Iqbal Medical College, Lahore, Pakistan, and included assessment of medical history, physical
10 examination, neurological evaluation. Brain MRI scans performed in Pakistan were analyzed in the
11 Neurobehavioral Clinical Research Section, National Human Genome Research Institute.

12 Identification of variants in *ZBTB20*

13 The pedigree of Family PKST77 is shown in Figure 1A. This family displays highly significant
14 genetic linkage between persistent stuttering and markers on chromosome 3q [8]. We re-investigated the
15 previously reported linkage using a genome-wide SNP linkage scan under a model of recessive
16 inheritance with reduced penetrance [8], which confirmed significant linkage scores in this region and
17 nowhere else across the genome. As previously observed with lower density markers, assignment of
18 precise linkage boundaries was limited by instances of apparent non-Mendelian segregation, consistent
19 with the accepted nature of stuttering as a complex trait. Whole-genome sequencing was performed in
20 16 individuals (III-13, III-17, III-21, III-22, IV-23, IV-24, IV-25, IV-26, IV-27, IV-28, IV-29, IV-30, IV-
21 31, IV-32, IV-33, IV-34) within this family. The whole-genome sequencing quality control followed the
22 minimum standards requirements established by The American Genome Center at Uniformed Service
23 University, with at least a minimum 30X Illumina coverage, 75% Q30 scores, and 100 gigabases of
24 sequence produced per sample. The analysis identified a mean of 3,824,970 SNVs (range 3,745,519 –
25 3,932,504 variants) and a mean of 1,023,407.56 indels (range 998,694 – 1,051,753 indels) (Suppl.1.1).

1 Once annotated, all variants underwent a bioinformatics sorting analysis as follows. First, variants
2 not on chromosome 3 were eliminated, leaving 12,283 variants (SNPs = 10,181; Indels = 2,102). Next,
3 variants residing outside the region of significant linkage (11,925 variants) were excluded, leaving a total
4 of 358 variants, all residing within the previously demonstrated linkage region. Many of these occurred
5 commonly (allele frequency ≥ 0.01) in gnomAD v2.1 (Genome Aggregation Database), the 1000
6 Genomes phase 3 dataset, or in the NHLBI ESP6400 dataset. These were eliminated from consideration,
7 and variants with minor allele frequency (MAF) < 0.01 in at least one of these databases were considered
8 further. Of the remaining 41 variants, we focused on the 17 coding variants (15 nonsynonymous, 1
9 frameshift deletion, and 1 stop gain, Table 1), based on the hypothesis that the causative allele in PKST77
10 could be a coding allele of large effect. Only one of these 17 rare variants co-segregated with the disorder
11 under the expected recessive model observed within the family. The variant (c.2155G>A) was in *ZBTB20*
12 (GenBank: NM_001164342). It occurs at the beginning of exon 5 and is predicted to cause a valine-to-
13 isoleucine substitution (p.Val719Ile), and the location is illustrated in Figure 2. The wild-type valine is
14 conserved across all vertebrates with the exception of coelacanth (*Latimeria chalumnae*), which carries a
15 phenylamine at this position (Figure 1C and Suppl.1.2). The Val719Ile change is predicted to be disease
16 causing (score:29) by Mutation Taster (<http://www.mutationtaster.org/>); probably damaging (score:
17 0.999) by PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>); and tolerated (score: 0.14) by SIFT
18 (<https://sift.bii.a-star.edu.sg/>).

19 In all family members previously analyzed with whole genome sequencing, the presence of the
20 homozygous variant c.2155G>A in *ZBTB20* was confirmed by Sanger sequencing as shown in Figure 1B.
21 In addition, we performed dideoxy sequencing in five additional unaffected members. Across the
22 individuals with fully confirmed genotype data, homozygosity for the Val719Ile variant co-segregates
23 with persistent stuttering.

24 This mutation is rare. It is present in 22/141,456 individuals in the gnomAD Database (v.2.1.1),
25 and all 22 of these individuals are heterozygous. We note that the individuals in the gnomAD database are
26 of unknown speech phenotype

1 (https://gnomad.broadinstitute.org/variant/rs779910215?dataset=gnomad_r2_1). A total of 13 other
2 missense variants were found in gnomAD, all present as heterozygotes. This indicates that the affected
3 individuals in PKST77 are the only known individuals homozygous for a coding mutation in *ZBTB20*.
4 We also analyzed our custom-built Communication Exome Database (CED) (N = 1,941 unrelated
5 individuals), sorting for coding variants among stuttering individuals in different cohort populations (N =
6 1,074 individuals) and matched controls (N = 867 individuals). No occurrences of the *ZBTB20*
7 c.2155G>A variant were identified.

8 To estimate the contribution of mutations in *ZBTB20* to stuttering overall, we performed a burden
9 test that compared the rate of rare variants in affected and control individuals. Although the frequency of
10 *ZBTB20* coding variants in our combined case groups (66/1,074) was nominally higher than that in the
11 combined controls (41/867), the difference in frequency between cases and controls did not reach
12 statistical significance in any of our population-specific case/control subgroups (Suppl.1.3). We also
13 compared the frequency of rare coding variants in our stuttering cohort (2,148 chromosomes) with that in
14 the general population represented in the gnomAD v.2.1 dataset, (282,912 chromosomes), which likely
15 contains both unaffected and affected individuals ([https://gnomad.broadinstitute.org/faq#what-](https://gnomad.broadinstitute.org/faq#what-populations-are-represented-in-the-gnomad-data)
16 [populations-are-represented-in-the-gnomad-data](https://gnomad.broadinstitute.org/faq#what-populations-are-represented-in-the-gnomad-data)). Here, the frequency of rare *ZBTB20* variants in our
17 stuttering group was higher overall. We observed more variants in total in our combined stuttering group,
18 which is highly ethnically diverse (88 alleles/2,148 total alleles) than in an ethnically diverse gnomAD
19 population containing similar proportions of European, Asian, and African individuals (5,450
20 alleles/282,912 total alleles, Chi-square = 52,716, $p = 3.85 \times 10^{-13}$). These data indicate that rare coding
21 variants of all types in *ZBTB20* are elevated in the stuttering populations.

22
23 The *ZBTB20*_Val719Ile mutation suppresses transcriptional repressor activity of *ZBTB20* in a human
24 cell.

25
26 *ZBTB20* has been reported to act as a transcriptional repressor for multiple target genes including
27 alpha fetoprotein (AFP) and inhibitor of kB alpha ($I\kappa B\alpha$) [26]. To investigate the functional effects of the

1 ZBTB20_Val719Ile mutation on protein activity in human cells, we performed a reporter gene assay
2 using the co-expression of ZBTB20 and luciferase proteins transcribed under control of the AFP
3 promoter. The wildtype and Val719Ile variants of ZBTB20 were transiently over-expressed separately in
4 the human HepG2 cells, along with AFP-luciferase vector, where the measured luciferase activity directly
5 reflected the intracellular activity of ZBTB as a transcriptional repressor. The luciferase activity was
6 reduced about 40% in both the low (20 ng) and high (40 ng) dosage of wildtype ZBTB20 transfection
7 comparing transfection to a control vector. Transfection of the same amount of ZBTB20_Ile719
8 expression vector rendered the transcriptional repressor activity significantly suppressed compared to
9 transfection of wildtype ZBTB20 in both low ($p = 0.034$) and high dosage ($p = 0.0083$). The data shown
10 in Figure 3 demonstrates that the ZBTB20_Ile719 is associated with a decrease in the activity of the
11 mutant ZBTB20 protein as a transcriptional repressor in human cells.

12 Ultrasonic mouse vocalizations

13 Animal models of stuttering have been created by generating mouse lines carrying a range of
14 mutations in the *GNPTAB* gene found in humans who stutter [12, 13]. Such animals displayed altered
15 ultrasonic vocalizations with abnormalities similar in some aspects to those observed in humans who
16 stutter. We used CRISPR-cas to engineer a mouse line carrying the Val719Ile mutation in *Zbtb20* [27,
17 Suppl.2.4]. Mice carrying the *Zbtb20* Val719Ile variant displayed normal pup isolation call vocalization
18 with respect to amount of vocalization and the spectral frequencies observed. The length of the silent
19 intervals in both the inter-bout and intra-bout vocalizations showed a suggestion of increased length
20 compared to those in their genetically wild-type littermates, but this difference was not statistically
21 significant (data not shown).

22

23 Clinical evaluations

24 *De novo* heterozygous missense variants in *ZBTB20* have previously been associated with
25 Primrose syndrome, a rare autosomal dominant disorder that is characterized by a slowly progressive
26 condition associated with tall stature, increased head circumference, atypical facial features, cognitive

1 deficits associated with autism spectrum disorder, and ectopic calcifications [28]. Primrose syndrome
2 patients may subsequently begin to manifest other clinical features including distal muscle atrophy,
3 hearing loss, cataracts, sparse body hair, disturbed glucose metabolism, dense calcification of external
4 ears, and anatomic abnormalities of the corpus callosum. Although Primrose syndrome is clinically
5 consolidated and recognizable in adults, the diagnosis in infants and children is considered difficult [29].
6 Four affected members of family PKST77 were available ten years after their original enrollment, and
7 they were examined in greater detail (Tables 2 and 3, Suppl.1.4). Many of the most common features of
8 Primrose syndrome, including tall stature and increased head circumference, were absent in these family
9 members. However mild facial dysmorphism and, in some members, mild intellectual disability was
10 noted.

11 Brain MRI scans were available for three family members, one homozygous normal
12 unaffected individual and two homozygous affected individuals. T1-weighted images (176 axial
13 slices, 0.9-mm slice thickness, 256×256 acquisition matrix) and diffusion weighted images
14 (TE=89ms, TR=11.1s, 128x128 matrix, slice thickness = 2 mm, gap = 2 mm, 102 directions)
15 were acquired in a 3-T Siemens MAGNETOM Vida scanner. Freesurfer version 7.1.1
16 (<https://surfer.nmr.mgh.harvard.edu>) and FSL version 6.0.1 (<https://fsl.fmrib.ox.ac.uk/>) were
17 used to process T1w and DWI data, respectively. Structural characteristics are shown in Table 2.
18 While this sample size was too small to draw statistically significant conclusions, no striking
19 differences were observed in the structure of the corpus callosum in these individuals.

20 In conclusion, the clinical phenotype information available from the members of PKST77 was
21 limited in our follow-up evaluations. In these evaluations, we noted some individuals with mild facial
22 dysmorphism and intellectual disability, although these were not consistent across affected individuals.
23 Both of these signs are observed in Primrose Syndrome, but our limited data precludes a firm conclusion
24 that this represents a variant form of this disorder.

25 **Discussion**

1 Persistent stuttering has been shown to be highly genetic in origin, and several individual
2 genes that contribute to this disorder have been described [10,11]. However, the genes found
3 thus far appear to account for only a fraction of cases of stuttering, and it is clear that additional
4 genetic factors exist. Data supporting the view that variants in *ZBTB20* cause of stuttering comes from
5 four lines of evidence. First, strong evidence for linkage of stuttering to this region has been documented
6 in a large, consanguineous family [8]. Second, homozygous mutations at a highly conserved site in the
7 protein encoded by this gene co-segregate with stuttering under an appropriate model of recessive
8 transmission with reduced penetrance, and no other variants in the linkage region passed our filtering
9 criteria. Third, in a population-based burden test we found a higher frequency of rare variants in this gene
10 in stuttering cases compared to ethnically matched control database groups. Fourth, this gene encodes a
11 transcription factor. The variant in this gene present in the large consanguineous family causes a
12 functional defect in its transcriptional regulatory activity in an *in vitro* luciferase reporter system. In
13 addition to these four lines of evidence, *ZBTB20* is known to be a major determinant of astrocyte
14 development in the brain [30, 31]. Studies have identified prominent astrocyte deficits in mice that
15 exhibit stuttering-like vocalization deficits due to a mutation in *GNPTAB*, a previously identified
16 stuttering gene [13]. Thus, our findings in *ZBTB20* may further support a role for astrocyte deficits in
17 humans who stutter.

18 Primrose syndrome is a dominant disorder, and all patients described to date carry a
19 mutation in just one of their two *ZBTB20* alleles. It is not yet known if these mutations act as
20 dominant negatives or via haploinsufficiency. Given the involvement of *ZBTB20* in a range of
21 important developmental processes [30, 31, 32, 33, 34], it is possible that homozygosity for
22 mutations in this gene is generally incompatible with viability. This idea is supported by the lack
23 of individuals homozygous for any coding sequence mutation in *ZBTB20* in the existing
24 gnomAD v.2.1 dataset, representing 141,456 multi-ethnic individuals.

1 We note that, in contrast to the homozygosity for the variant in family PKST77, our population-
2 based burden test identified mostly heterozygous *ZBTB20* variants. This was the case for both our
3 stuttering groups and control groups. We suggest that while the unusual homozygosity in family PKST77
4 resulted in identification of a genotype with large effect, other *ZBTB20* alleles may exert an effect in a
5 heterozygous state, but discernable only at the population level. It is also noteworthy that stuttering has
6 not been described in Primrose Syndrome to date (<https://www.omim.org/entry/259050>). We note that
7 the stuttering genes identified thus far had been previously associated with other rare Mendelian medical
8 genetic disorders, none of which had stuttering as a noted feature of their clinical presentation [10, 11,
9 <https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Fact-Sheets/Mucopolipidoses-Fact-Sheet>,
10 <https://omim.org/entry/607244>]. The reasons for this are not fully understood, but for one of these genes,
11 it has been shown that the mutational spectrum present in stuttering does not overlap with the mutations
12 in this gene seen in the rare medical genetic disorder [35].

13 *Zbtb20* is fundamental in brain development and function, and it is expressed in
14 developing hippocampal neurons, as well as in astrocytes in the cerebral cortex and cerebellum
15 [32, 33, 34] where it has been demonstrated to be an essential regulator of astrocyte
16 development. During this development, *Zbtb20* is highly expressed in neural precursor cells at
17 late stages of neocortical development, and it is involved in the differentiation and maturation of
18 astrocytes while suppressing the development of neurons [30].

19 Animal models of stuttering have been created by inserting human stuttering mutations
20 into the mouse [12, 13] and such mice display abnormalities in their ultrasonic vocalizations that
21 share similarities with the speech deficits in human who stutter and carry these same mutations.
22 The mice display anomalies in their brain astrocytes, and this pathology is significantly displayed
23 in the corpus callosum [13]. Structural abnormalities of the corpus callosum are a feature of
24 Primrose syndrome [14, 20, 36], and although we did not observe such abnormalities in the

1 members of family PKST077, the small number of subjects available for evaluation limits this
2 conclusion. It is interesting to note that this gene plays a major role in the development of mammalian
3 astrocytes [30], a cell type previously discovered in a mouse model of stuttering caused by another gene
4 [13]. While our findings do not speak to the state of potential astrocyte pathology in the members of
5 family PKST77, they suggest a functional deficit in the transcriptional regulatory function of ZBTB20 in
6 this family. If astrocyte pathology is associated with human stuttering, our findings here would be
7 consistent with this, and should warrant additional investigation of this so-called astrocyte hypothesis in
8 human stuttering. We note these findings implicate astrocyte pathology indirectly, and that previous
9 findings were found with mutations in a different gene (*GNPTAB*) using a mouse model of stuttering.
10 The present findings in human stuttering, originally based on a mechanistically agnostic linkage study,
11 demonstrate an alteration in function in a transcription factor necessary for astrocyte development.

12 We suggest that mutations in *ZBTB20* are a relatively rare cause of stuttering, as such
13 mutations were observed in only a small number of unrelated individuals who stutter beyond
14 family PKST077. Nevertheless, we believe these findings open new avenues of research into the
15 neuropathological underpinnings of this enigmatic disorder.

16 **Materials and Methods**

17 **Subjects**

18 All human subjects, including individuals from a consanguineous family (PKST77), with
19 the inclusion diagnosis of non-syndromic persistent stuttering were enrolled with written
20 informed consent under NIH protocol 97-DC-0057 (Genetic studies of stuttering), also reviewed
21 and approved by the IRB of the Centre of Excellence in Molecular Biology, University of the
22 Punjab, as previously described (Raza et al., 2010). All subjects were age 8 and older.
23 Documented neurologically normal control DNA samples (NDPT; N = 460) were obtained from
24 the National Institute of Neurological Disorders and Stroke (NINDS) panels NDPT006,

1 NDPT020, NDPT023, NDPT079, NDPT082, NDPT093 and NDPT094 from the Coriell Cell
2 Repository (available at
3 [https://www.coriell.org/Search?grid=0&q=ninds&csId=&f_0=NINDS+Repository&f_1=Control](https://www.coriell.org/Search?grid=0&q=ninds&csId=&f_0=NINDS+Repository&f_1=Control&f_2=96+well+plate+of+DNA+samples)
4 [&f_2=96+well+plate+of+DNA+samples](https://www.coriell.org/Search?grid=0&q=ninds&csId=&f_0=NINDS+Repository&f_1=Control&f_2=96+well+plate+of+DNA+samples)). Our study also included unrelated affected individuals
5 (total N = 1,074) and population-matched control individuals (N = 867), which was made up of
6 Pakistani affected individuals (PKSTR; N = 72) and Pakistani normally fluent controls (PKNR;
7 N = 96), Cameroonian affected (STCR; N = 105) and control (RC; N = 93) individuals, North
8 American affected individuals, including those from our NIH group (NIH; N = 739) and
9 Brazilian affected individuals (BRMII; N = 158) and controls (BRCO; N = 218). Stuttering was
10 diagnosed as previously described [37, 38]. Individuals were classified as affected as previously
11 described [10]. Affected individuals were classified as unrelated by self-report, and no genotypic
12 evidence for the relatedness among these subjects was observed. The identity of all subjects is
13 protected under the privacy and confidentiality guidelines of the National Institutes of Health
14 Institutional Review Board (Protocol 97-DC-0057), and the identities of the subjects are not
15 known outside of the research group.

16 Linkage analysis

17 Two independent linkage analyses were performed in family PKST77. The first was
18 described by Raza et al. (2010). The second, newly performed in this study, used the same
19 subjects, here genotyped at 551,839 SNPs assayed on the Illumina Infinium CoreExome-24 v.1.1
20 arrays according to manufacturer's protocols. The linkage analysis was performed using
21 MERLIN 1.1.2. (<https://csg.sph.umich.edu/abecasis/Merlin/download/>) with parametric and non-
22 parametric models.

23 Whole Exome and Genomic Sequencing

1 Genomic DNA was purified from whole-blood samples using a PAXgene[®] Blood DNA
2 Kit (QIAGEN), or from saliva samples using an Oragene[®] DNA kit (DNA Genotek Inc.,
3 Canada), according to the manufacturer's protocols. Exome enriched DNA libraries were
4 prepared using Nextera Rapid Capture Exome (Illumina, San Diego, CA, USA) and were
5 sequenced on an Illumina HiSeq1500 with 75 x 75 paired-end reads in 1,941 samples.
6 Additionally, whole-genome libraries were prepared on 16 members of the family PKST77 using
7 an Illumina Nextera TruSeq DNA PCR-free kit and sequenced on an Illumina HiSeq X Ten at
8 The American Genome Center at Uniformed Service University (TAGC-USU,
9 <https://www.usuhs.edu/chirp/bioinformatics>).

10

11 Sequence Data Analyses

12 Whole-exome and whole-genome sequence quality control reads were aligned to Genome
13 Reference Consortium Human Build 38 patch release 12 (GRCh38.p12).

14 *Bioinformatic pipelines*

15 Whole-exome data was analyzed using the bcbio-nextgen germline variant pipeline
16 (<https://bcbio-nextgen.readthedocs.io/en/latest/>). For variant calling, the “ensemble” vcf file was
17 used, which required that a variant be detected in at least two of the following variant callers:
18 gatk-haplotype, platypus, vascan, freebayes, or Samtools. The complete snakemake pipeline
19 script is provided in the supplement (Suppl.2.1 and Suppl.2.2).

20 Whole-genome data: for genomic alignment was used NovoAlign - Novocraft 3.08.02
21 (<http://www.novocraft.com/products/novoalign/>), followed by Picard Toolkit 2.9.2
22 (<https://broadinstitute.github.io/picard/>), the variants were called using Genome Analysis Tool
23 Kit 3.8.0 (GATK) (<https://software.broadinstitute.org/gatk/>) according to Best Practices

1 Workflow (<https://software.broadinstitute.org/gatk/best-practices/>) and, called variants were
2 annotated using ANNOVAR, release 2017Jul16.
3 (<http://annovar.openbioinformatics.org/en/latest/>).

4

5 Database annotation and variant sort

6 A total of 1,941 samples among unrelated stuttering cases and controls were analyzed in
7 a custom-built Communication Exome Database (CED). This MySQL relational database
8 included metadata on each individual, variants from the corresponding “ensembl.vcf” file, and
9 extensive metadata on each variant including annotations from the Genome Aggregation
10 Database (gnomAD v2.1) (<https://gnomad.broadinstitute.org/>), 1000 Genomes Project phase 3
11 (http://useast.ensembl.org/Homo_sapiens/Info/Index), National Heart, Lung, and Blood Institute
12 – Exome Sequencing Project (NHLBI-ESP) Exome Variant Server
13 (<https://evs.gs.washington.edu/EVS/>), Kaviar Project ([http://db.systemsbiology.net/kaviar/cgi-](http://db.systemsbiology.net/kaviar/cgi-pub/Kaviar.pl)
14 [pub/Kaviar.pl](http://db.systemsbiology.net/kaviar/cgi-pub/Kaviar.pl)), dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>). Variants with minor allele
15 frequency ≥ 0.01 were filtered out. The remaining variants, sorted according to chromosome
16 position, gene function, type of amino acid changed, evolutionary conservation level and clinical
17 information using annotations provided by public databases (including RefSeq
18 [<https://www.ncbi.nlm.nih.gov/refseq/>], knownGene [<http://genome.ucsc.edu/>] dbNSFP
19 database, [<https://sites.google.com/site/jpopgen/dbNSFP>] and ClinVar database
20 [<https://www.ncbi.nlm.nih.gov/clinvar/>]).

21 Sanger sequencing

22 Sanger sequencing was performed to validate all variants previously identified by whole
23 exome sequencing. Oligos were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and

1 are listed in the supplemental materials (Suppl.2.3). Sanger dideoxy sequencing of all exons and
2 their 40 bp flanking regions was performed using BigDye[®] Terminator Cycle Sequencing Kit
3 v3.1 (Applied Biosystems, Foster City, CA, USA) with standard manufacturer's protocols.
4 Sequence reactions were resolved on an ABI Genetic Analyzer 3730xl instrument (Applied
5 Biosystems) and the electropherograms were evaluated using DNASTAR[®] Lasergene SeqMan
6 Pro[™] (v.13.0.2) (<https://www.dnastar.com/software/lasergene/>).

7

8 Construction of *ZBTB20* expression vectors

9 *ZBTB20* expression vector was prepared using OriGene (OriGene Tech. Inc., MD),
10 TrueORF[®] cDNA clone (Cat#: RC228567) (available in: [https://www.origene.com/catalog/cdna-](https://www.origene.com/catalog/cdna-clones/expression-plasmids/rc228567/zbtb20-nm_001164342-human-tagged-orf-clone)
11 [clones/expression-plasmids/rc228567/zbtb20-nm_001164342-human-tagged-orf-clone](https://www.origene.com/catalog/cdna-clones/expression-plasmids/rc228567/zbtb20-nm_001164342-human-tagged-orf-clone)). This
12 vector expresses *ZBTB20* human isoform 1 (NM_001164342) under the control of a CMV
13 promoter. The wild type valine at position 719 of *ZBTB20* was replaced with isoleucine by
14 introducing a point mutation using QuickChange II Site-Directed Mutagenesis kit (Agilent Tech.
15 Inc., CA). The promoter of alpha fetoprotein (*AFP*, -2320 to 23 bp) was amplified from genomic
16 DNA of a normal human individual using a KOD Hot Start DNA Polymerase (EMD Millipore
17 Corp., US) with primers of 5'-AAGGTACCTTACTCCTGGTGTAGTCGCATC -3' and 5'-
18 AAAACTCGAGTGGCAGTGGTGGGAAGCACAATA -3', which were designed to introduce
19 KpnI and XhoI restriction enzyme sites (underlined) at the 5' and 3' ends of the promoter
20 sequence, respectively. These restriction sites were used in subcloning of the PCR product into
21 pGL4.10[luc2] Firefly Luciferase Vector (Promega Corp., WI).

22

23 Luciferase reporter assay

1 In 96-well white solid plates microplates, 4×10^4 HepG2 cells were seeded and
2 transiently transfected next day with 100 ng of plasmids which include 40 ng of *ZBTB20*-
3 expression vector (pCMV6-*ZBTB20*) or empty expression vector (pCMV6-Entry Tagged
4 Cloning Vector, OriGene Tech. Inc., MD) (Cat#: PS100001), 40 ng of pGL4.10-*AFP*, and 20 ng
5 of *Renilla* Luciferase Control Reporter Vector as a transfection efficiency control (pRL-TK,
6 Promega Corp., WI) using Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher
7 Scientific Inc., CA) according to the manufacturer's instructions. Opti-MEM®| Reduced-Serum
8 Medium (Gibco™) was replaced by Eagle's Minimum Essential Medium (EMEM) (ATCC®)
9 with 10% fetal bovine serum after 4 hours of transfection. Luciferase activity was measured
10 using the Dual-Glo® Luciferase Assay System (Promega Corp., WI) after 27 hours of
11 transfection with Mithras LB940, Multimode Microplate Reader (Berthold Tech. GmbH & Co.,
12 Germany). All experiments were performed with four trials, with four replicates in each trial.

13

14 Clinical Studies

15 Clinical and Neurological evaluations of the members of family PKST077 were
16 performed at Allama Iqbal Medical Research Institute, Lahore, Pakistan and at the National
17 Human Genome Research Institute, NIH. Brain MRI studies were available from individuals IV-
18 24, IV-26, and IV-27 (Figure 1A).

19 2.10. KI mouse model

20 We used the CRISPR/Cas9 genome editing technology [27] to generate a knock-in mouse
21 model that mimics the mutation founded in the affected members of stuttering family PKST77.
22 (<http://www.informatics.jax.org/marker/MGI:1929213>), several different crRNAs were designed

1 to target the sequence near to the variant and minimize potential off-target sites. More details are
2 available in the Supplementary material 2.4 (Suppl.2.4).

3

4 Ultrasound mouse vocalization analysis

5 The ultrasonic vocalizations of mice carrying the equivalent of the human Ile719Val
6 mutation and their wild type littermates were gathered as previously described [12].

7 **Abbreviations used**

8 NIDCD- National Institute on Deafness and Other Communication Disorders

9 ZBTB20- Zinc finger and BTB domain containing, 20

10 NHLBI- National Heart Lung and Blood Institute

11 CED- Communications Exome Database

12 AFP- Alpha fetoprotein

13 I κ B α - Inhibitor of κ B alpha

14 GNPTAB- N-acetylglucosamine-1-phosphate transferase subunits alpha and beta

15 CRISPR-cas - Clustered Regularly Interspaced Short Palindromic Repeats-Crispr-associated
16 protein 9

17 IRB- Institutional Review Board

18 NINDS- National Institute of Neurological Diseases and Stroke

19 PKSTR- Pakistani stuttering affected individuals, unrelated

20 PKNR- Pakistani normally fluent controls, unrelated

21 STCR- Cameroonian stuttering affected individuals, unrelated

22 RC- Cameroonian normal control individuals, unrelated

23 BRMII- Brazilian stuttering affected individuals, unrelated

1 BRCO- Brazilian normally fluent control individuals, unrelated

2 EMEM- Eagle's Minimum Essential Medium

3 KI- Knock-in

4

5 **Declarations**

6 **Ethical approval and consent to participate** – Subjects were enrolled with written informed
7 consent under NIH Institutional Review Board-approved Consent # 97-DC-0057 entitled Genetic
8 Studies of Stuttering.

9 **Consent for publication** – Subjects consented to publication without identifying information
10 under NIH Institutional Review Board-approved Consent # 97-DC-0057 entitled Genetic Studies
11 of Stuttering.

12 **Availability of Supporting data** – Supporting data not presented in the Supplementary materials
13 is available from the communicating author at: drayna@nidcd.nih.gov

14 **Competing Interests** - The authors declare no competing interests.

15 **Funding** - This work was supported in part by the National Institute on Deafness and Other
16 Communication Disorders (NIDCD) Intramural grant # Z01-DC-000047-19, the NIDCD
17 Genomic and Computational Biology Core # Z1C-000086,

18 **Author contributions** - CFD performed DNA sequencing, and linkage and bioinformatic
19 analyses, MHR ascertained and sampled subjects and performed linkage analyses and DNA
20 sequencing, TUH performed luciferase reporter studies, TB performed mouse ultrasonic
21 vocalization studies, PS and GS evaluated MRI and other clinical data, SR ascertained subjects
22 and supervised research in Pakistan, RM oversaw whole exome and whole genome sequencing,
23 and DD conceived of the study, obtained funding, oversaw human subjects protection, and

1 supervised all aspects of research. All authors contributed to writing and reviewing the
 2 manuscript.

3 **Acknowledgements** - This work used the high performance computational capabilities of the
 4 Biowulf Linux Cluster at the National Institutes of Health, Bethesda, MD
 5 (<http://biowulf.nih.gov>). We thank Dr. Timothy Holy of Washington University St. Louis for
 6 support of TB, and The Stuttering Foundation, the National Stuttering Association, and Hollins
 7 Communications Research Institute, Roanoke, VA for assistance with subject recruitment.

8 **Authors' information:**

9 Carlos E Frigerio Domingues¹⁺, Muhammad Hashim Raza^{1+^}, Tae-Un Han^{1#}, Terra Barnes², Philip Shaw³,
 10 Gustavo Sudre³, Sheikh Riazuddin⁴, Robert J. Morell⁵, Dennis Drayna^{1*}

11

12 1. Laboratory of Communication Disorders, National Institute on Deafness and Other Communication
 13 Disorders, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

14 2. Department of Neuroscience, School of Medicine, Washington University in St. Louis, MO

15 3. Neurobehavioral Clinical Research Section, National Human Genome Research Institute, National
 16 Institutes of Health, Bldg. 31, Bethesda MD 20892

17 4. Allama Iqbal Medical College, University of Health Sciences, Lahore, Pakistan

18 5. Genomics and Computational Biology Core, National Institute on Deafness and Other Communication
 19 Disorders, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

20 + These authors contributed equally to this research

21 # Current address: Molecular Neurogenetics Section, National Human Genome Research Institute,
 22 National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892

23 % Current address: Utah Center for Genetic Discovery, Eccles Institute of Human Genetics, University of
 24 Utah, Salt Lake City, UT 84112

25 * To whom correspondence should be addressed at: NIDCD/National Institutes of Health
 26 Porter Neuroscience Research Center
 27 35A Convent Drive, Bethesda, MD 20892
 28 Tel. 301-221-4392
 29 drayna@nidcd.nih.gov
 30 ORCID 0000-0002-1730-4532

31

32

1 References

- 2 1. American Psychiatric Association & Force, D. S. M. T. (2013). *Diagnostic and statistical manual of*
3 *mental disorders : DSM-5*. American Psychiatric Association, Washington, D.C.
- 4 2. International Classification of Diseases, 10th Edition (ICD-10), World Health Organization (2013)
- 5 3. Frigerio-Domingues, C., & Drayna, D. (2017). Genetic contributions to stuttering: the current evidence.
6 *Mol Genet Genomic Med*, 5(2), 95-102. doi:10.1002/mgg3.276
- 7 4. Kidd, K. K., Heimbuch, R. C., & Records, M. A. (1981). Vertical transmission of susceptibility to
8 *stuttering with sex-modified expression*. *Proc Natl Acad Sci U S A*, 78(1), 606-610.
- 9 5. Domingues, C. E., Olivera, C. M., Oliveira, B. V., Juste, F. S., Andrade, C. R., Giacheti, C. M., . . . Drayna,
10 D. (2014). A genetic linkage study in Brazil identifies a new locus for persistent developmental
11 *stuttering on chromosome 10*. *Genet Mol Res*, 13(1), 2094-2101. doi:10.4238/2014.March.24.13
- 12 6. Raza, M. H., Amjad, R., Riazuddin, S., & Drayna, D. (2012). Studies in a consanguineous family reveal a
13 *novel locus for stuttering on chromosome 16q*. *Hum Genet*, 131(2), 311-313.
14 doi:10.1007/s00439-011-1134-2
- 15 7. Raza, M. H., Gertz, E. M., Mundorff, J., Lukong, J., Kuster, J., Schaffer, A. A., & Drayna, D. (2013).
16 *Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance*.
17 *Hum Genet*, 132(4), 385-396. doi:10.1007/s00439-012-1252-5
- 18 8. Raza, M. H., Riazuddin, S., & Drayna, D. (2010). Identification of an autosomal recessive stuttering
19 *locus on chromosome 3q13.2-3q13.33*. *Hum Genet*, 128(4), 461-463. doi:10.1007/s00439-010-
20 0871-y
- 21 9. Riaz, N., Steinberg, S., Ahmad, J., Pluzhnikov, A., Riazuddin, S., Cox, N. J., & Drayna, D. (2005).
22 *Genomewide significant linkage to stuttering on chromosome 12*. *Am J Hum Genet*, 76(4), 647-
23 651. doi:10.1086/429226
- 24 10. Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J. C., & Drayna, D. (2010).
25 *Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering*. *N Engl J Med*,
26 362(8), 677-685. doi:10.1056/NEJMoa0902630
- 27 11. Raza, M. H., Mattera, R., Morell, R., Sainz, E., Rahn, R., Gutierrez, J., . . . Drayna, D. (2015a).
28 *Association between Rare Variants in AP4E1, a Component of Intracellular Trafficking, and*
29 *Persistent Stuttering*. *Am J Hum Genet*, 97(5), 715-725. doi:10.1016/j.ajhg.2015.10.007
- 30 12. Barnes, Terra D., Wozniak, David F., Gutierrez, J., Han, T.-U., Drayna, D., & Holy, Timothy E. (2016). A
31 *Mutation Associated with Stuttering Alters Mouse Pup Ultrasonic Vocalizations*. *Current Biology*,
32 26(8), 1009-1018. doi:<http://dx.doi.org/10.1016/j.cub.2016.02.068>
- 33 13. Han, T. U., Root, J., Reyes, L. D., Huchinson, E. B., Hoffmann, J. D., Lee, W. S., . . . Drayna, D. (2019).
34 *Human GNPTAB stuttering mutations engineered into mice cause vocalization deficits and*
35 *astrocyte pathology in the corpus callosum*. *Proc Natl Acad Sci U S A*, 116(35), 17515-17524.
36 doi:10.1073/pnas.1901480116
- 37 14. Alby, C., Boutaud, L., Bessières, B., Serre, V., Rio, M., Cormier-Daire, V., . . . Thomas, S. (2018). Novel
38 *de novo ZBTB20 mutations in three cases with Primrose syndrome and constant corpus callosum*
39 *anomalies*. *Am J Med Genet A*, 176(5), 1091-1098. doi:10.1002/ajmg.a.38684
- 40 15. Casertano, A., Fontana, P., Hennekam, R. C., Tartaglia, M., Genesio, R., Dieber, T. B., . . . Melis, D.
41 (2017). *Alterations in metabolic patterns have a key role in diagnosis and progression of*
42 *primrose syndrome*. *Am J Med Genet A*, 173(7), 1896-1902. doi:10.1002/ajmg.a.38124
- 43 16. Cleaver, R., Berg, J., Craft, E., Foster, A., Gibbons, R. J., Hobson, E., . . . Tatton-Brown, K. (2019).
44 *Refining the Primrose syndrome phenotype: A study of five patients with ZBTB20 de novo*
45 *variants and a review of the literature*. *Am J Med Genet A*, 179(3), 344-349.
46 doi:10.1002/ajmg.a.6102491

- 1 17. Cordeddu, V., Redeker, B., Stellacci, E., Jongejan, A., Fragale, A., Bradley, T. E., . . . Hennekam, R. C.
2 (2014). Mutations in ZBTB20 cause Primrose syndrome. *Nat Genet*, 46(8), 815-817.
3 doi:10.1038/ng.3035
- 4 18. Ferreira, L. D., Borges-Medeiros, R. L., Thies, J., Schnur, R. E., Lam, C., & de Oliveira, J. R. M. (2019).
5 Expansion of the Primrose syndrome phenotype through the comparative analysis of two new
6 case reports with ZBTB20 variants. *Am J Med Genet A*, 179(11), 2228-2232.
7 doi:10.1002/ajmg.a.61297
- 8 19. Grímsdóttir, S., Hove, H. B., Kreiborg, S., Ek, J., Johansen, A., Darvann, T. A., & Hermann, N. V. (2019).
9 Novel de novo mutation in ZBTB20 in primrose syndrome in boy with short stature. *Clin*
10 *Dysmorphol*, 28(1), 41-45. doi:10.1097/mcd.0000000000000244
- 11 20. Mattioli, F., Piton, A., Gérard, B., Superti-Furga, A., Mandel, J. L., & Unger, S. (2016). Novel de novo
12 mutations in ZBTB20 in Primrose syndrome with congenital hypothyroidism. *Am J Med Genet A*,
13 170(6), 1626-1629. doi:10.1002/ajmg.a.37645
- 14 21. Posmyk, R., Leśniewicz, R., Chorży, M., & Wołczyński, S. (2011). New case of Primrose syndrome
15 with mild intellectual disability. *Am J Med Genet A*, 155a(11), 2838-2840.
16 doi:10.1002/ajmg.a.34257
- 17 22. Stellacci, E., Steindl, K., Joset, P., Mercurio, L., Anselmi, M., Cecchetti, S., . . . Rauch, A. (2018). Clinical
18 and functional characterization of two novel ZBTB20 mutations causing Primrose syndrome.
19 *Hum Mutat*, 39(7), 959-964. doi:10.1002/humu.23546
- 20 23. Yamamoto-Shimojima, K., Imaizumi, T., Akagawa, H., Kanno, H., & Yamamoto, T. (2020). Primrose
21 syndrome associated with unclassified immunodeficiency and a novel ZBTB20 mutation. *Am J*
22 *Med Genet A*, 182(3), 521-526. doi:10.1002/ajmg.a.61432
- 23 24. Zhang, W., Mi, J., Li, N., Sui, L., Wan, T., Zhang, J., . . . Cao, X. (2001). Identification and
24 characterization of DPZF, a novel human BTB/POZ zinc finger protein sharing homology to BCL-6.
25 *Biochem Biophys Res Commun*, 282(4), 1067-1073. doi:10.1006/bbrc.2001.4689
- 26 25. Mitchelmore, C., Kjaerulff, K. M., Pedersen, H. C., Nielsen, J. V., Rasmussen, T. E., Fisker, M. F., . . .
27 Jensen, N. A. (2002). Characterization of two novel nuclear BTB/POZ domain zinc finger isoforms.
28 Association with differentiation of hippocampal neurons, cerebellar granule cells, and macroglia.
29 *J Biol Chem*, 277(9), 7598-7609. doi:10.1074/jbc.M110023200
- 30 26. Xie, Z., Zhang, H., Tsai, W., Zhang, Y., Du, Y., Zhong, J., . . . Zhang, W. J. (2008). Zinc finger protein
31 ZBTB20 is a key repressor of alpha-fetoprotein gene transcription in liver. *Proc Natl Acad Sci U S*
32 *A*, 105(31), 10859-10864. doi:10.1073/pnas.0800647105
- 33 27. Henao-Mejia, J., Williams, A., Rongvaux, A., Stein, J., Hughes, C., & Flavell, R. A. (2016). Generation of
34 Genetically Modified Mice Using the CRISPR–Cas9 Genome-Editing System. 2016(2),
35 *pdb.prot090704*. doi:10.1101/pdb.prot090704
- 36 28. Primrose, D. A. (1982). A slowly progressive degenerative condition characterized by mental
37 deficiency, wasting of limb musculature and bone abnormalities, including ossification of the
38 pinnae. *J Ment Defic Res*, 26 (Pt 2), 101-106. doi:10.1111/j.1365-2788.1982.tb00133.x
- 39 29. Melis, D., Carvalho, D., Barbaro-Dieber, T., Espay, A. J., Gambello, M. J., Gener, B., . . . Hennekam, R.
40 C. (2020). Primrose syndrome: Characterization of the phenotype in 42 patients. *Clin Genet*,
41 97(6), 890-901. doi:10.1111/cge.13749
- 42 30. Nagao, M., Ogata, T., Sawada, Y., & Gotoh, Y. (2016). *Zbtb20* promotes astrocytogenesis during
43 neocortical development. *Nat Commun*, 7, 11102. doi:10.1038/ncomms11102
- 44 31. Medeiros de Araújo J, Barão S, Mateos-White I, Espinosa A, Costa MR, Gil-Sanz C, Müller U. ZBTB20
45 is crucial for the specification of a subset of callosal projection neurons and astrocytes in the mammalian
46 neocortex. *Development* 2021 Aug 15;148(16):dev196642. doi:10.1242/dev.196642. Epub 2021 Aug 19.

- 1 32. Nielsen, J. V., Blom, J. B., Noraberg, J., & Jensen, N. A. (2010). *Zbtb20-induced CA1 pyramidal neuron*
 2 *development and area enlargement in the cerebral midline cortex of mice. Cereb Cortex, 20(8),*
 3 *1904-1914. doi:10.1093/cercor/bhp261*
- 4 33. Nielsen, J. V., Thomassen, M., Møllgård, K., Noraberg, J., & Jensen, N. A. (2014). *Zbtb20 defines a*
 5 *hippocampal neuronal identity through direct repression of genes that control projection neuron*
 6 *development in the isocortex. Cereb Cortex, 24(5), 1216-1229. doi:10.1093*
- 7 34. Sakaguchi, H., Kadoshima, T., Soen, M., Narii, N., Ishida, Y., Ohgushi, M., . . . Sasai, Y. (2015).
 8 *Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-*
 9 *derived dorsomedial telencephalic tissue. Nat Commun, 6(1), 8896. doi:10.1038/ncomms9896*
- 10 35. Raza M, Domingues C, Webster R, Sainz E, Paris E, Rahn, R, Gutierrez J, Chow HM,
 11 Mundorff J, Kang C, Riaz N, Basra M, Khan S, Riazuddin S, Moretti-Ferreira D,
 12 Braun A, and Drayna D. *Mucopolidosis Type II/III alpha and non-syndromic stuttering are*
 13 *associated with different variants in the same genes. European Journal of Human Genetics*
 14 *24(4):529-536 (2015b)*
- 15 36. Juven, A., Nambot, S., Piton, A., Jean-Marçais, N., Masurel, A., Callier, P., . . . Faivre, L. (2020).
 16 *Primrose syndrome: a phenotypic comparison of patients with a ZBTB20 missense variant versus*
 17 *a 3q13.31 microdeletion including ZBTB20. Eur J Hum Genet, 28(8), 1044-1055.*
 18 *doi:10.1038/s41431-020-0582-3*
- 19 37. Riley, G. D. (1994). *Stuttering Severity Instrument for Children and Adults - (SSI-3) (3th ed. ed.).*
 20 *Austin: Pro Ed.*
- 21 38. Webster, R. L. (1980). *Evolution of a target-based behavioral therapy for stuttering. J Fluency Disord,*
 22 *5(3), 303-320. doi:[http://dx.doi.org/10.1016/0094-730X\(80\)90035-2](http://dx.doi.org/10.1016/0094-730X(80)90035-2)*

25 Competing Interests

26 The authors have no relevant financial or non-financial interests to disclose.

27 Data Availability

28 The datasets generated during and/or analysed during the current study are available from the
 29 corresponding author on reasonable request.

31 Figure Captions

32 **Figure 1 A.** Pedigree of stuttering family PKST77. Individuals diagnosed as affected represented by filled
 33 symbols. wt, wildtype allele; m, mutated allele; wt/wt, homozygote wildtype; wt/m, heterozygote
 34 mutated; m/m, homozygote mutated **B.** Dideoxy Sanger sequencing traces showing the c.2155G>A
 35 variant in *ZBTB20* in available family members. **C.** Amino acid alignments across species of the Valine
 36 719, in bold

37 **Figure 2** Chromosome 3 linkage region, structure of *ZBTB20* gene, and mutation location

38 **Figure 3** The *ZBTB20_Val719Ile* variant decreases transcriptional repressor activity. Human HepG2 cells
 39 were co-transfected with the pCMV6-ZBTB20 wild type or Val719Ile mutation, and the pGL4.10-AFP
 40 reporter vector expressing a luciferase gene under control of the AFP promoter (-2320 to +23). A Renilla-
 41 luciferase vector (pRL-TK) was also co-transfected to control for transfection efficiency. Relative
 42 expression level was calculated dividing luciferase activity of pCMV6-ZBTB20 by that of pCMV6_Empty

- 1 vector as a mock control. A total of 16 replications in three independent experiments are shown (total n
- 2 = 16, 16). Legend: Wt, wildtype ZBTB20; Error bar shows SEM. * p value = 0.034; ** p value = 0.0083

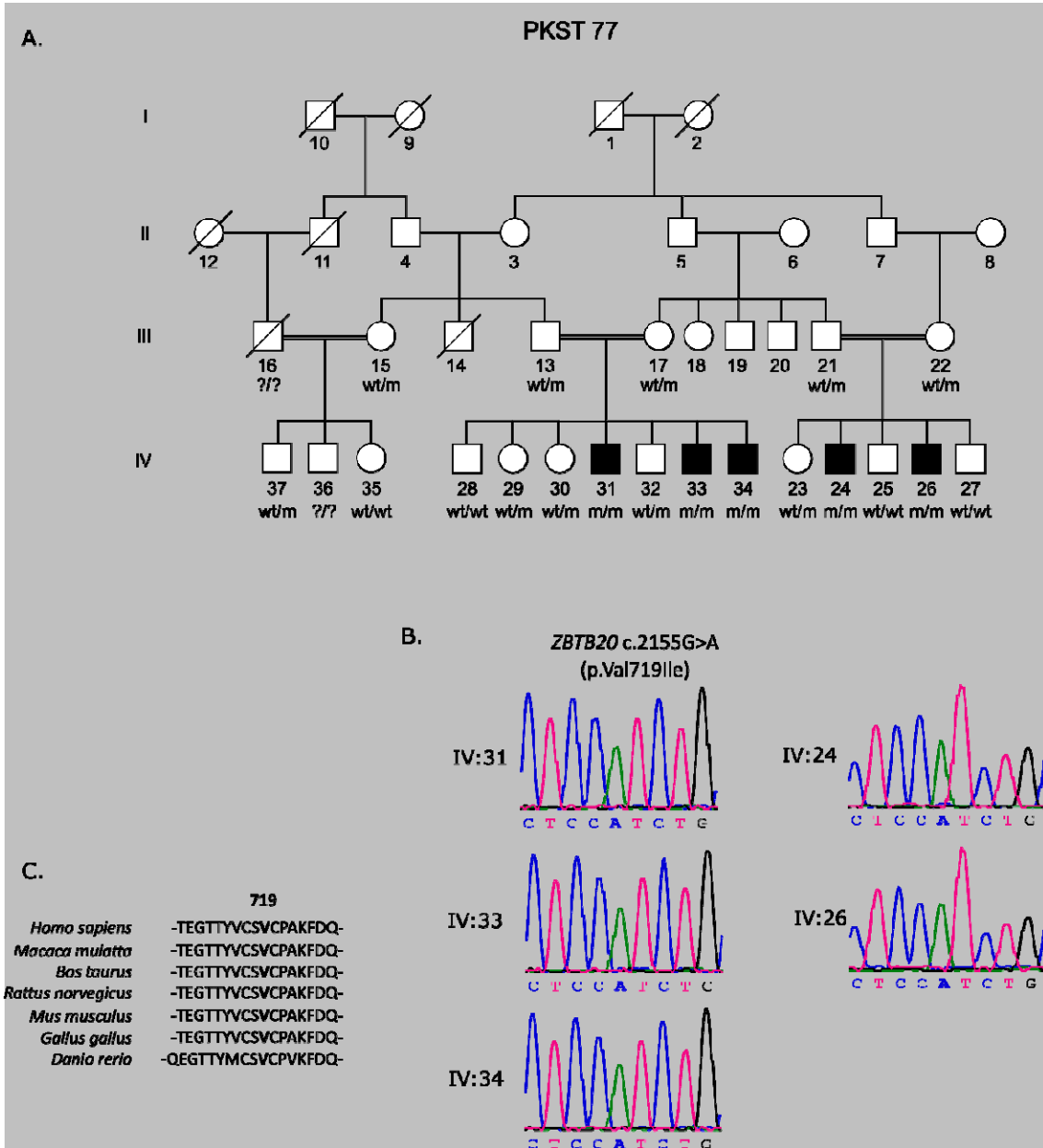


Figure 2.

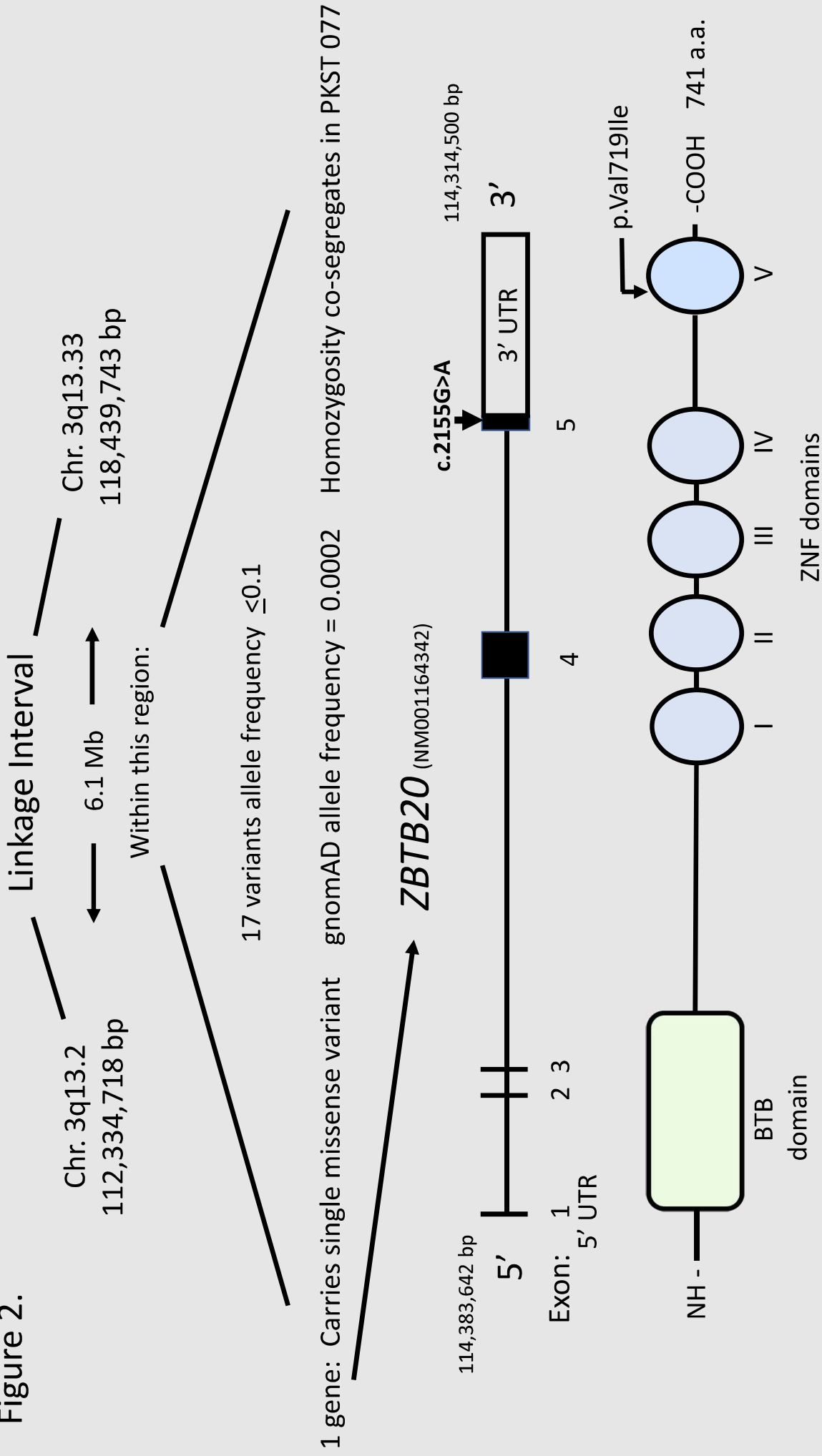
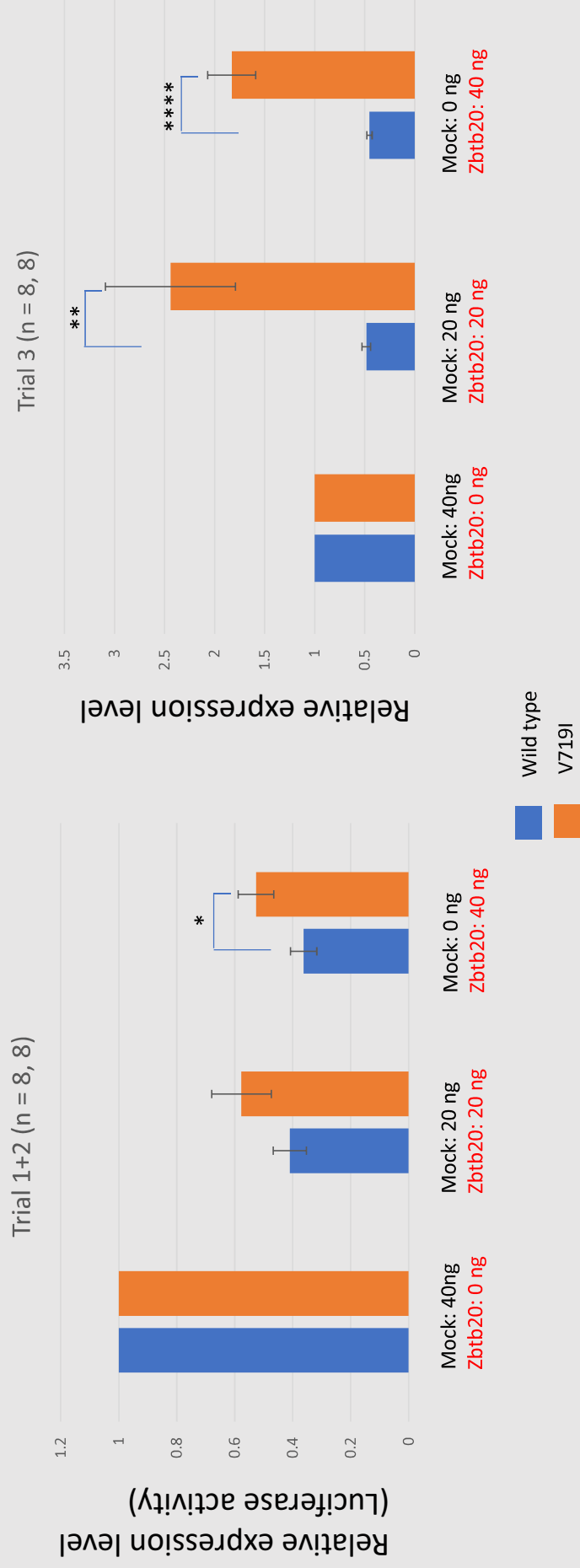


Figure 3.

Target: Afp Cell: HepG2



Zbtb20_V719I variant decreases transcriptional repressor activity

Chr	Start	End	rsID	Ref	Alt	Region	Gene name	Type	Exon	c.DNA position	Amino Acid change
3	112113049	112113049	rs191148598	G	A	exonic	<i>C3orf52</i>	nonsynonymous	Exon 5	c.553G>A	p.E185K
3	112974799	112974801	rs564728039	AAG	-	exonic	<i>CD200R1</i>	nonframeshift deletion	Exon 1	c.57_59del	p.19_20del
3	112991054	112991054	rs565265086	G	A	exonic	<i>GTPBP8</i>	nonsynonymous	Exon 1	c.55G>A	p.V19M
3	114339076	114339076	rs779910215	C	T	exonic	<i>ZBTB20</i>	nonsynonymous	Exon 5	c.2155G>A*	p.V719I
3	119034641	119034641	rs576681103	G	C	UTR5	<i>IGSF11</i>	-	-	-	-
3	119147246	119147246	rs554707851	C	G	exonic	<i>C3orf30</i>	nonsynonymous	Exon 1	c.1057C>G	p.Q353E
3	119732456	119732456	rs773076124	A	G	exonic	<i>MAATS1</i>	nonsynonymous	Exon 9	c.1181A>G	p.Y394C
3	120348820	120348820	rs201718706	G	T	exonic	<i>LRRC58</i>	nonsynonymous	Exon 1	c.424C>A	p.L142M
3	121476641	121476641	rs not annotated	G	C	exonic	<i>POLQ</i>	nonsynonymous	Exon 20	c.6304C>G	p.Q2102E
3	121923043	121923043	rs753517272	C	T	exonic;splicing	<i>SLC15A2</i>	stopgain	Exon 9	c.778C>T	p.Q260X
3	122568928	122568928	rs150765058	C	T	exonic	<i>DTX3L</i>	nonsynonymous	Exon 3	c.839C>T	p.T280I
3	123448538	123448538	rs1279127415	C	A	exonic	<i>ADCY5</i>	nonsynonymous	Exon 1	c.8G>T	p.G3V
3	123700313	123700313	Rs1452968448	T	C	exonic	<i>MYLK</i>	nonsynonymous	Exon 15	c.3155A>G	p.N1052S
3	125091448	125091448	rs762284041	G	A	exonic	<i>SLC12A8</i>	nonsynonymous	Exon 6	c.1315C>T	p.L439F
3	127003384	127003384	rs199693063	G	A	exonic	<i>PLXNA1</i>	nonsynonymous	Exon 3	c.1432G>A	p.V478I
3	128910310	128910310	rs765706722	G	A	exonic	<i>ACAD9</i>	nonsynonymous	Exon 8	c.1192C>T	p.P398S
3	128945350	128945350	rs775891329	C	T	exonic	<i>CFAP92</i>	nonsynonymous	Exon 3	c.617G>A	p.R206H

Table 1. List of 17 coding variants (MAF < 0.01) under final investigation.

Note: **ZBTB20* variant present in all affected individuals in homozygous state

Table 2. PKST 077 Clinical evaluation report.

Clinical finds	Patient ID			
	PKST77-33	PKST77-34	PKST77-24	PKST77-26
Gender	Male	Male	Male	Male
Weight (kg)	56	58	50	50
Head circumference (cm)	55.5	56.0	49.5	48.0
Height (meters)	1.75	1.75	1.52	1.52
Developmental Milestones Delayed*				
Copy Sounds (within six months)	+	+	+	-
Sit without support (within six months)	+	+	+	-
Like to play with others, especially parents (within six months)	+	+	+	-
Respond to own name (within six months)	+	+	+	-
Copy gestures (6-12 months)	+	+	+	-
Respond to simple spoken requests (6-12 months)	+	+	+	-
Say “mama” and “dada” (6-12 months)	+	+	+	-
Pull up to stand (6-12 months)	+	+	+	-
Say several single words (12-18 months)	+	+	+	-
Walk alone (12-18 months)	+	+	< 2 years	4 years
Say sentence with 2 to 4 words (18-24 months)	+	+	+	-
Follow simple instructions (18-24 months)	+	+	+	-
Point to things or pictures when they are named (18-24 months)	+	+	+	-
Carry on a conversation using 2-3 sentences (24-36 months)	+	+	+	-
Climb well (24-36 months)	+	+	+	-
Copy friend and adults (like running when other children run) (24-36 months)	+	+	+	-
Hop and stand on one foot for up to 2 seconds (36-48 months)	+	+	+	-
Tell stories (36-48 months)	+	+	+	-
Draw a person with 2 to 4 body parts (36-48 months)	+	+	+	-
Morphology, neuromuscular and system involvement findings				
Is there history of birth asphyxia?	-	-	-	-
Skin pigmentation abnormalities	-	-	-	-

Table 2. PKST 077 Clinical evaluation report. (*continuation*)

Clinical finds	Patient ID			
	PKST77-33	PKST77-34	PKST77-24	PKST77-26
Morphology, neuromuscular and system involvement findings				
Facial dysmorphic features	+	+	+	+ ¹
Chin abnormalities	n.a.	-	-	Retrognathia
Motor weakness	-	-	-	-
Chest abnormalities	-	-	Barrel shaped	-
Dental abnormalities	-	-	-	Dysplasia
Ear abnormalities	-	n.a.	-	-
Eye abnormalities	-	-	-	+ ²
Balance problems	-	-	-	-
Can the patient stand without swaying to any side with eyes open and closed?	-	+	+	+
Can patient walk on a straight line?	-	+	+	-
Mood	n.a.	n.a.	labile	n.a.
Family history				
Miscarriages	n.a.	n.a.	-	-
Intellectual disability	+	+	+	+
Congenital abnormalities	-	-	-	+

Notes: * Reference age in months; + present; - absent; n.a. not available; (1) Prominent supraorbital ridges, (2) Bilateral pterygium, left eye divergent squint

Table 3. Structural characteristics of the Corpus Callosum in members of Pakistani stuttering family PKST077.

PKST77 Family members	Corpus Callosum Volume (mm³)	Corpus Callosum Fraction Anisotropy (FA, n.u.)
PKST77-24	3652.1	0.36
PKST77-26	4994.3	0.32
PKST77-27	3982.8	0.34