The American Journal of Human Genetics, Volume *91*

Supplemental Data

Exome Sequencing Identifies *FUS* Mutations as a Cause of Essential Tremor

Nancy D. Merner, Simon L. Girard, Hélène Catoire, Cynthia V. Bourassa, Veronique V. Belzil, Jean-Baptiste Rivière, Pascale Hince, Annie Levert, Alexandre Dionne-Laporte, Dan Spiegelman, Anne Noreau, Sabrina Diab, Anna Szuto, Hélène Fournier, John Raelson, Majid Belouchi, Michel Panisset, Patrick Cossette, Nicolas Dupré, Geneviève Bernard, Sylvain Chouinard, Patrick A. Dion, and Guy A. Rouleau

Supplemental Subjects and Methods

Recruitment and Diagnosis

Ethics approval for the recruitment and genetic analysis of ET affected individuals and their families has been granted at the following institutes, the CRCHUM (Centre de recherche du Centre hospitalier de l'Université de Montréal) (project no: ND043076), the CHA (Centre hospitalier affilié universitaire de Québec) (project no: PEJ-280), and the Sainte-Justine University Hospital Center (project no: 2352). Recruitment has been ongoing over the last ten years; the Rouleau laboratory has recruited over 80 French Canadian probands with hereditary ET. Participating individuals gave written consent and allowed blood withdrawal for DNA extraction and cell line establishment.

ET has been diagnosed as either 'definite,' 'probable,' or 'possible' according to the criteria specified by the MDS ET consensus statement and Louis et al. in 1998[1; 2]. Exclusion criteria included if (i) an identified cause of exaggerated physiological tremor was noted, (ii) other neurological deficits (Parkinsonisms, polyneuritis, other) were present, and if (iii) an orthostatic tremor or (iv) a psychogenic-like tremor was observed.

The first and largest family to be recruited to the Rouleau laboratory, FET1, was selected for exome sequencing for this current study (Figure S1). Thirty-five individuals from FET1 were clinically observed (Figure S2) and willing to participate in the genetic study; 23 individuals were diagnosed as affected (7 'definitely' affected, 3 'probably' affected, and 13 'possibly' affected), and 12 individuals were noted to be clinically unaffected (Figure S1). This family was clinically assessed originally in 1998, and again in 2008.

Exome Capture and Sequencing

Exome sequencing was carried out on four individuals from FET1 with a 'definite' ET diagnosis and an age of onset under the age of 40 years, along with a clinically unaffected married-in family member (Figure S1). Targeted enrichment was performed using the Agilent SureSelect all exome kit-G3362D with 2 µg of genomic DNA. Exon-enriched DNA libraries from these five individuals were sequenced on the Applied Biosystems SOLiD^{TM} 4 System; four samples were run per slide.

Read Mapping

BFAST was used to align the sequence reads to the human genome (hg18)[3]; all parameters of BFAST were kept at the default settings and, as recommended by the authors, 10 genome indexes were used for optimize the alignment. All PCR duplicates were removed from the alignments and only uniquely mapped reads were considered. Coverage of the mapped reads to the exon targets was calculated by read depth at each nucleotide position that was sequenced.

All human exons represent approximately 38 Mb of the human genome; the actual number of base pairs covered by the Agilent exome probe kit was 38,801,372 (Table S1). In this study, the average number of base pairs that was covered greater than or equal to 1X equalled 33,386,274 (representing 86% of the Agilent kit), and 15X equalled 23,144,747 (representing 60% of the Agilent kit). The total number of uniquely mapped reads per individual averaged to be 26,023,097 (Figure S3 and Table S1).

Variant Calling and Annotation

A minimum coverage of 15X, a mutation frequency of 20% (with the mutation seen in at least two different reads), and a minimal sequence quality score (defined as the average Phred quality score for all reads at a particular variant position) of 20 were required for variants to be called. Varscan was used for the variant detection[4], and ANNOVAR was used for the annotation against the RefSeq database and dbSNP[5].

Segregation Analysis

After a list of variants was generated for each sequenced family member, a segregation analysis was performed to determine which variants were shared exclusively by affected individuals and not by the married-in control. An additional filter was set to determine which of those variants were detected in a group of 94 control exomes that were also sequenced in the Rouleau laboratory in order to remove (i) common sequencing artifacts and (ii) common variants with an allele frequency above 1%.

Validation of Exome Sequencing Variants through Sanger Sequencing

Primers were designed using Primer3[6] for exome sequencing variants that were identified only in 'definitely' affected individuals in FET1 and had a sequencing quality score greater than or equal to 50. Polymerase chain reactions (PCRs) were performed using the AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, California, USA) as per manufacturer's instruction. To visualize DNA fragments, 5 μL of the PCR product was loaded on a 1% agarose gel; ethidium bromide was used for the staining. PCR products were sequenced at the Genome Quebec Innovation Centre (Montréal, Québec, Canada) using a 3730XL DNAnalyzer (Applied Biosystems, Foster City, California, USA), and analyzed using the SoftGenetics program, Mutation surveyor (v.3.10, SoftGenetics, State College, Pennsylvania, USA).

Screening of a Control Cohort

Only one variant was validated through Sanger sequencing, which happened to be in the *FUS*/*TLS* (fused in sarcoma/translocated in liposarcoma) gene (accession number

NM_004960.3), thus previously published *FUS* screening data of a ethnically-matched control cohort from the Rouleau laboratory was used[7]. The allele frequency of ET *FUS* variants was determined using this control cohort.

Gene Validation

A cohort of 270 ET cases was screened for rare coding variants in the entire ORF (open reading frame) of *FUS*; primer sets were previously published[8]. The screening cohort included 61 familial cases from the Rouleau laboratory, and 96 familial and 113 sporadic cases from Genizon Biosciences Incorporated.

Protein Sequence Alignment and In Silico Prediction Programs

Conservation of the FUS protein across species was determined using Weblogo[9]; the following orthologues were aligned: *Homo sapiens* (NP_004951), *Bos taurus* (NP_776337), *Rattus norvegicus* (NP_001012137), *Mus musculus* (NP_631888), *Gallus gallus* (NP_001001531), *Xenopus laevis* (NP_001080383), and *Danio rerio* (NP_957377). The effects of amino acid substitutions on protein function were also predicted using MutationTaster, SIFT and Polyphen[10-12]. Variant splicing predictions were made using a splicing site prediction program by Neural Networks (part of the Berkeley Drosophila Genome Project) and MutationTaster[10; 13]. FUS domains were predicted using the following websites, http://www.uniprot.org/uniprot/P35637 and http://www.cbs.dtu.dk/services/NetNES/.

Cell Culture and Treatment with Puromycin

Lymphoblast lines were established from peripheral blood samples as described[14]. Cells were maintained in RPMI 1640 with 2 mM L-glutamine and 10% FBS in a 37^{o}C incubator (5% CO_2). Treatments with puromycin were carried out by incubating 10 million cells with 300 μg/ml of this drug (Sigma-Aldrich) for 6 hours.

RNA Extraction, cDNA Synthesis, and Splicing Experiments

Total RNA was extracted using Trizol (Invitrogen) from lymphoblastoid cells (puromycin treated and non-treated) prepared from five different *FUS* c.868C>T (p.Gln290Ter) mutation-positive individuals from FET1 (Individuals II:13, III:14, IV:6, IV:11, IV:12 in Figure S1), an ET-affected individual with *FUS* c.646C>T (p.Arg216Cys), an ET *FUS* c.1292C>T (p.Pro431Leu) mutation-positive individual, as well as, three ALS-affected individuals who each carried a different *FUS* mutation (either c.1555C>T (p.Gln519Ter), c.1562G>A (p.Arg521His), or the splice variant c.1542-2A>C), and two healthy controls. The QuantiTect® Reverse Transcription protocol from Qiagen was used for the synthesis of cDNA used in normal RT-PCR, and the Superscript® VILO™ cDNA Synthesis kit from Invitrogen for the cDNA used in quantitative RT-PCR; ratios (A260/A280) were all above 1.6. Primers for the splicing experiment were designed using Primer3 and *FUS* accession number NM_004960.3.

Primer Name	Primer Sequence	Amplicon Size
FUS-cDNA-5'UTR-ex5-F	Ccaggcgtcggtactcag	490
FUS-cDNA-5'UTR-ex5-R	Tttgctgctgtccaccatag	
FUS-cDNA-ex3-6-F	Acggacacttcaggctatgg	478
FUS-cDNA-ex3-6-R	Cacctccaccactctggtct	
FUS-cDNA-ex5-10-F	Ccccagagtgggagctacag	600
FUS-cDNA-ex5-10-R	Caaattaatcatgggctgtcc	
FUS-cDNA-ex6-12-F	Ggccaagatcaatcctccat	669
FUS-cDNA-ex6-12-R	Accacctccatagcctccac	
FUS-cDNA-ex9-13-F	Tggctgattacttcaagcaga	493
FUS-cDNA-ex9-13-R	Cccatgtgagagccacct	
FUS-cDNA-ex13-3'UTR-F	Gagggggaccaggtggctctcacatgggg	519
FUS-cDNA-ex13-3'UTR-R	Ggatatcatttggccttctccccgaac	

Quantitative RNA Expression Assays

Quantitative RT-PCR was performed using the TaqMan method (Applied Biosystems). Overall expression of *FUS* was performed using an inventoried set of probe and primers (Hs00192029_m1), and c.868 allele specific expression with a custom SNP genotyping set of 2 probes and primers designed by Applied Biosystems based on cDNA sequence (the normal allele, c.868C was detected with a VIC probe, and the mutated allele, c.868T with a FAM probe). PCR conditions were as follows: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles at 95°C for 15 sec (denaturation) and 60°C for 1 min (annealing and extension). Fluorescent signals were captured using the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems). The level of expression was determined by converting the threshold cycle (Ct) values using the $2^{-\Delta\Delta Ct}$ method. Expression levels were normalized using the human 18S ribosomal RNA (rRNA) gene with commercial primer-probe mix (Applied Biosystems) and were calculated in comparison to the average level of expression of two healthy controls. T-tests were used for statistics.

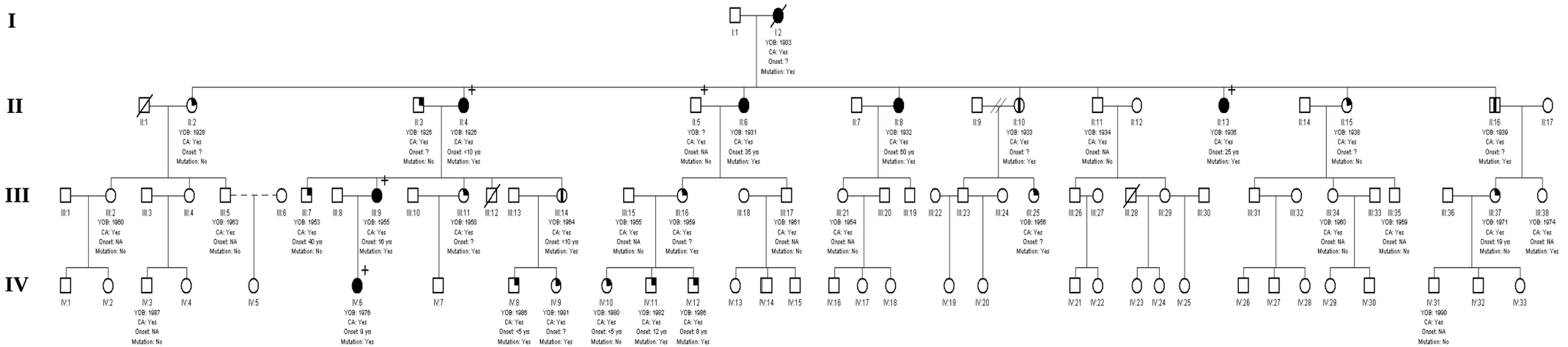

Figure S1. Essential Tremor Family FET1

Individuals with a definite diagnosis of ET are represented by symbols filled completely black. Individuals with a probable diagnosis of ET have symbols with a vertical black line in the center. Individuals with a possible diagnosis of ET have symbols with a small black square in the top right corner. Individuals that were used for exome sequencing are labelled with a "+" at the top right of the symbol. The year of birth (YOB), clinical assessment (CA), age of onset and mutation status are noted under each individual that participated in the study. "NA" means "not applicable."

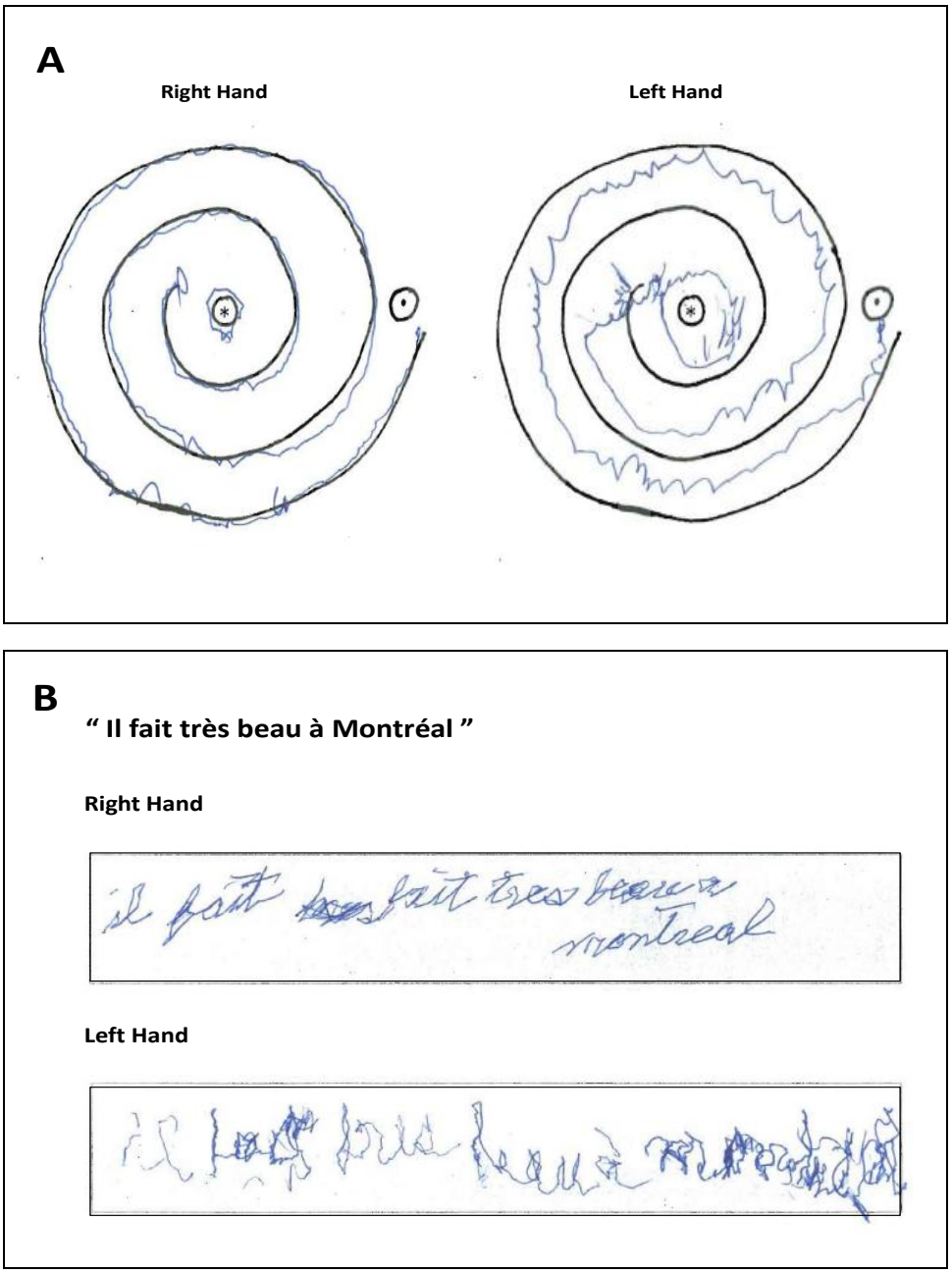

Figure S2. Writing Samples from the Definitely Affected Individual II:4

(A) The Archimedes Spiral Test - a widely used shape tracing task used to assess tremor. The individual was asked to make a spiral starting from the point in the middle, keep the pen between the lines and not to take the pen of the paper.

(B) The Writing Test - the individual was asked to write the sentence "Il fait très beau à Montréal."

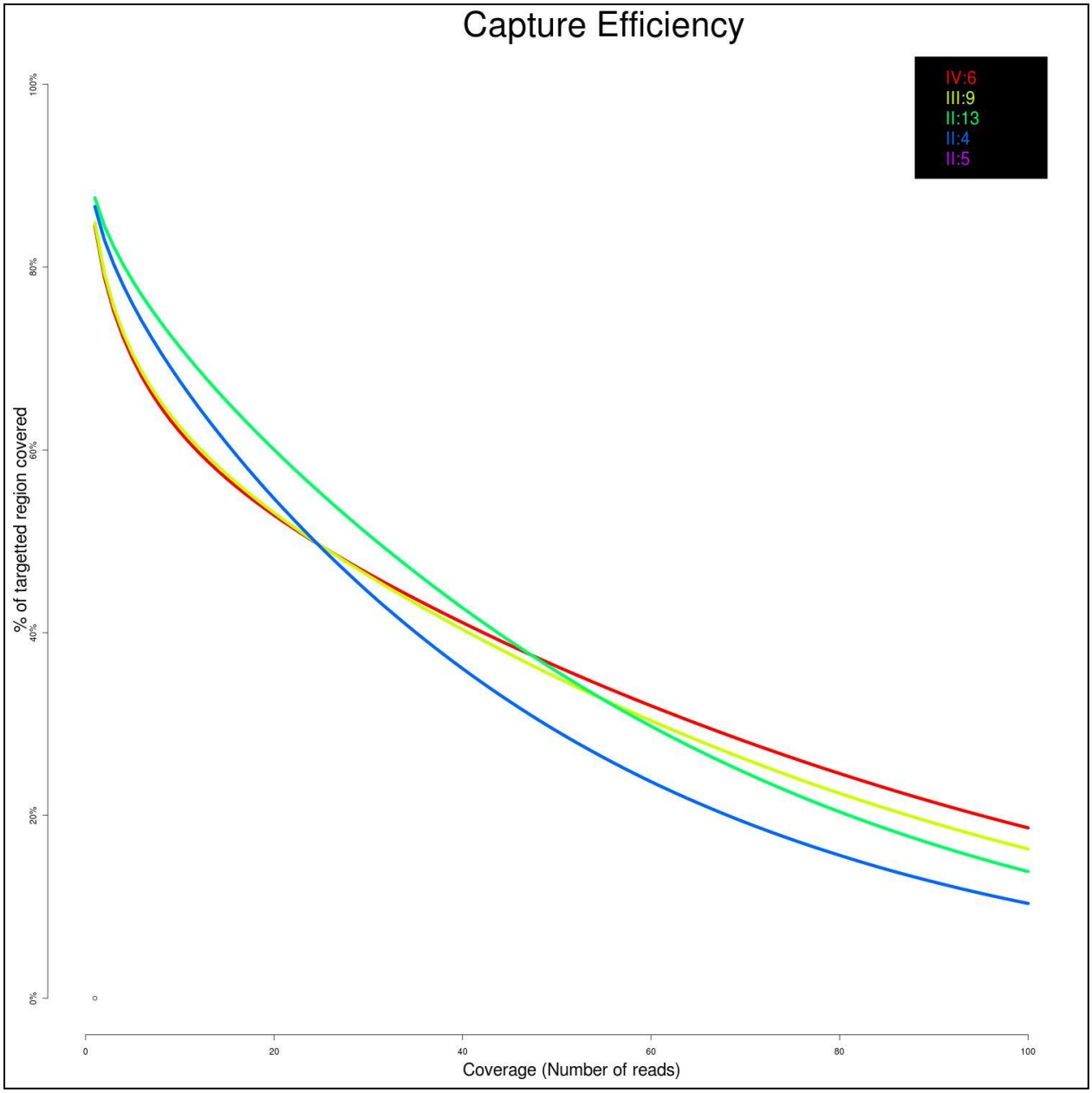

Figure S3. Exome Capture Efficiency

The exome capture efficiency is presented for each individual sequenced in family FET1. The x-axis represents the coverage in total number of reads, and the y-axis is the percentage of the total targeted region, on a per-bases calculation.

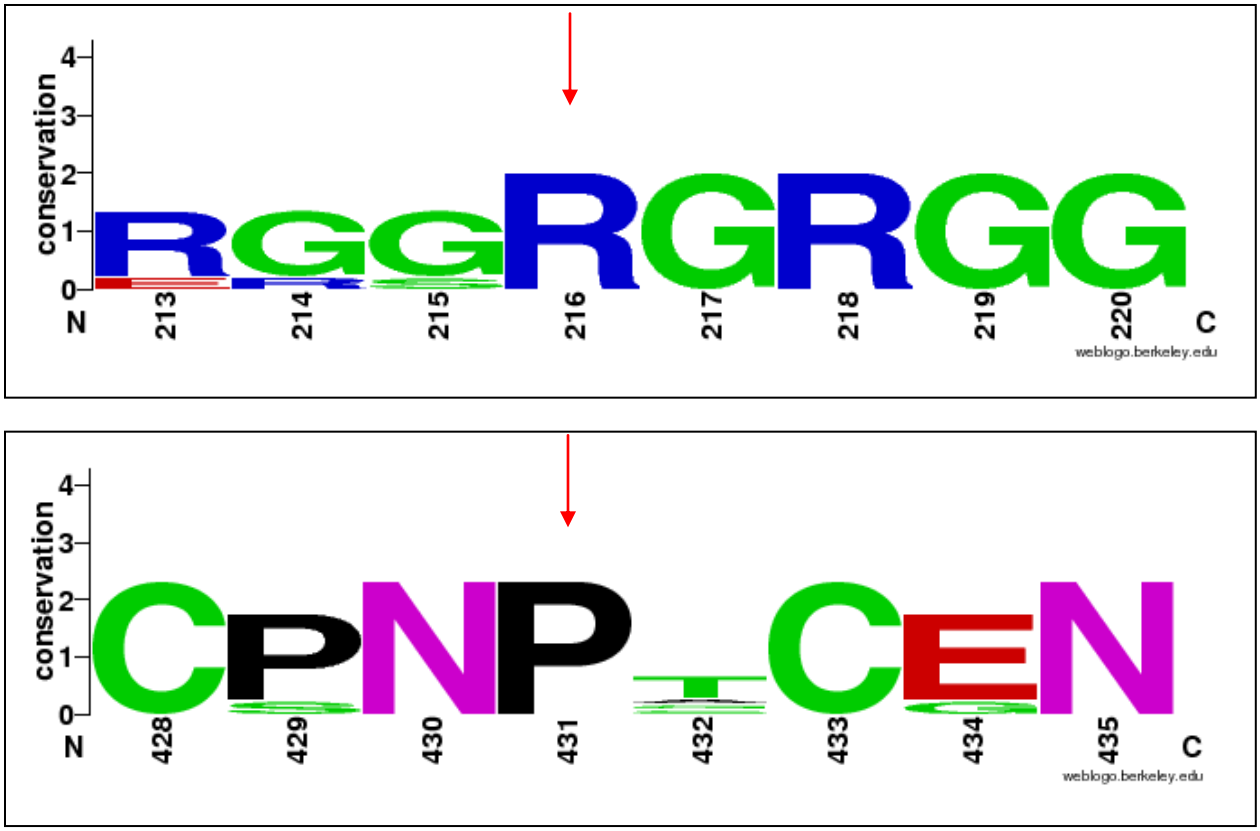

Figure S4. The Conservation of FUS Amino Acids p.Arg216 and p.Pro431

The following orthologues were aligned using Weblogo: *Homo sapiens* (NP_004951), *Bos taurus* (NP_776337), *Rattus norvegicus* (NP_001012137), *Mus musculus* (NP_631888), *Gallus gallus* (NP_001001531), *Xenopus laevis* (NP_001080383), and *Danio rerio* (NP_957377). The y-axis in this figure serves as a means of determining relative conservation and is not an actual measurement: the height of each stack at each amino acid position is relative to the overall conservation at that position, the height of the letters within each stack indicate the relative frequencies for each amino acid possibility, and the width of each stack corresponds to the proportion of valid readings at that position (indicating if sequence gaps exist between the shown amino acids).

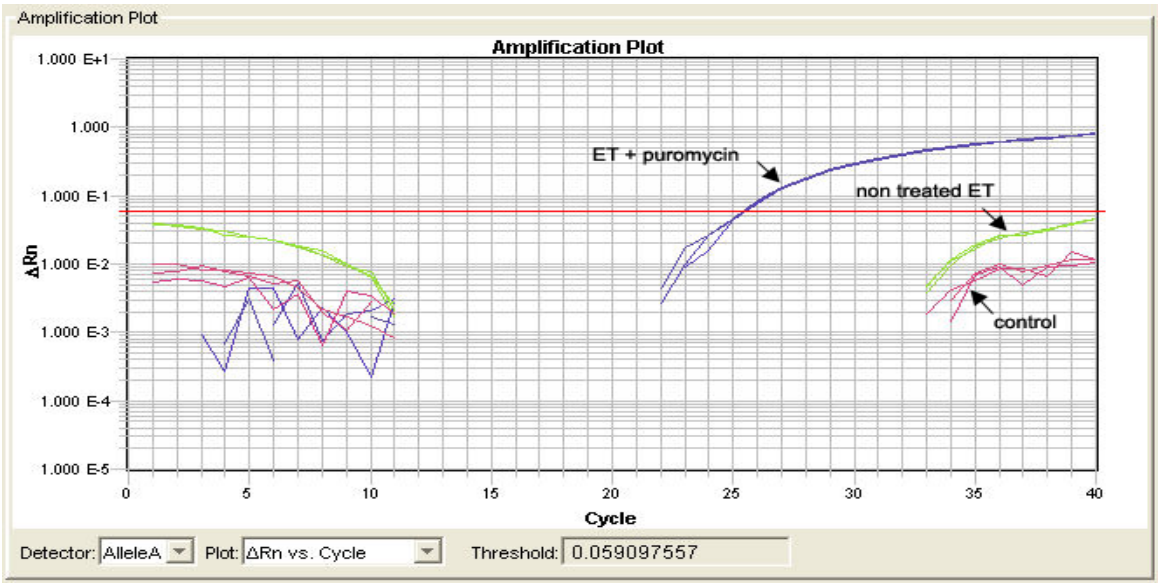

Figure S5. Raw Data of Quantitative *FUS* c.868C>T Allele Specific Expression

Raw data curves from the results of quantitative RT-PCR reactions in Figure 2C. The red line indicates the automatic threshold over which quantification is best measured. The dark blue measures are from puromycin treated cells from an ET affected individual; green from an untreated ET affected individual's cells; dark pink from a healthy individual with no *FUS* c.868C>T variation.

Table S1. Exome-Sequencing Statistics for FET1 Family Members

FET1		Number of base pairs targeted by the Agilent exome probe kit	Base pairs covered greater than or equal to 1X		Base pairs covered greater than or equal to 15X		Number of uniquely mapped reads
Individual	Clinical Status		Number	Percentage in Agilent probe kit	Number	Percentage in Agilent probe kit	
IV:6	affected	38,801,372	33,185,397	86%	22,247,102	57%	25,563,340
III:9	affected		32,904,599	85%	22,125,103	57%	28,622,732
II:13	affected		33,031,081	85%	22,308,341	57%	26,741,248
II:4	affected		33,719,715	87%	23,624,633	61%	22,655,426
II:5	control		34,090,579	88%	25,418,558	66%	26,532,741
Average	-	38,801,372	33,386,274	86%	23,144,747	60%	26,023,097

Table S2. *FUS* c.868C>T Mutation Status in FET1 Family Members

Diagnosis	c.868C>T carrier			Non-carrier			Total number of individuals	Total number of mutation carriers	Percentage of mutation carriers
	Under 40 yrs	Over or equal to 40 yrs	Unknown age	Under 40 yrs	Over or equal to 40 yrs	Unknown age			
Definite	5	1	1	0	0	0	7	7	100%
Probable	1	0	2	0	0	0	3	3	100%
Possible	3	0	4	2	1	3	13	7	54%
Clinically unaffected	-	-	-	-	-	-	12	1	8%

All individuals that were clinically observed are categorized on the basis of clinical status, age of onset, and mutation status.

Supplemental References

1. Louis, E.D., Ford, B., Lee, H., Andrews, H., and Cameron, G. (1998). Diagnostic criteria for essential tremor: a population perspective. Arch Neurol 55, 823-828.
2. Deuschl, G., Bain, P., and Brin, M. (1998). Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. Mov Disord 13 Suppl 3, 2-23.
3. Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. PLoS One 4, e7767.
4. Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25, 2283-2285.
5. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38, e164.
6. Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132, 365-386.
7. Belzil, V.V., Daoud, H., St-Onge, J., Desjarlais, A., Bouchard, J.P., Dupre, N., Lacomblez, L., Salachas, F., Pradat, P.F., Meininger, V., et al. (2011). Identification of novel FUS mutations in sporadic cases of amyotrophic lateral sclerosis. Amyotroph Lateral Scler 12, 113-117.
8. Belzil, V.V., Valdmanis, P.N., Dion, P.A., Daoud, H., Kabashi, E., Noreau, A., Gauthier, J., Hince, P., Desjarlais, A., Bouchard, J.P., et al. (2009). Mutations in FUS cause FALS and SALS in French and French Canadian populations. Neurology 73, 1176-1179.
9. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res 14, 1188-1190.
10. Schwarz, J.M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7, 575-576.
11. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 3812-3814.
12. Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30, 3894-3900.
13. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J Comput Biol 4, 311-323.
14. Anderson, M.A., and Gusella, J.F. (1984). Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. In Vitro 20, 856-858.