**The American Journal of Human Genetics, Volume *101***

## Supplemental Data

# A Pentanucleotide ATTTC Repeat Insertion in the Non-coding Region of *DAB1*, Mapping to *SCA37*, Causes Spinocerebellar Ataxia

**Ana I. Seixas, Joana R. Loureiro, Cristina Costa, Andrés Ordóñez-Ugalde, Hugo Marcelino, Cláudia L. Oliveira, José L. Loureiro, Ashutosh Dhingra, Eva Brandão, Vitor T. Cruz, Angela Timóteo, Beatriz Quintáns, Guy A. Rouleau, Patrizia Rizzu, Ángel Carracedo, José Bessa, Peter Heutink, Jorge Sequeiros, Maria J. Sobrido, Paula Coutinho, and Isabel Silveira**

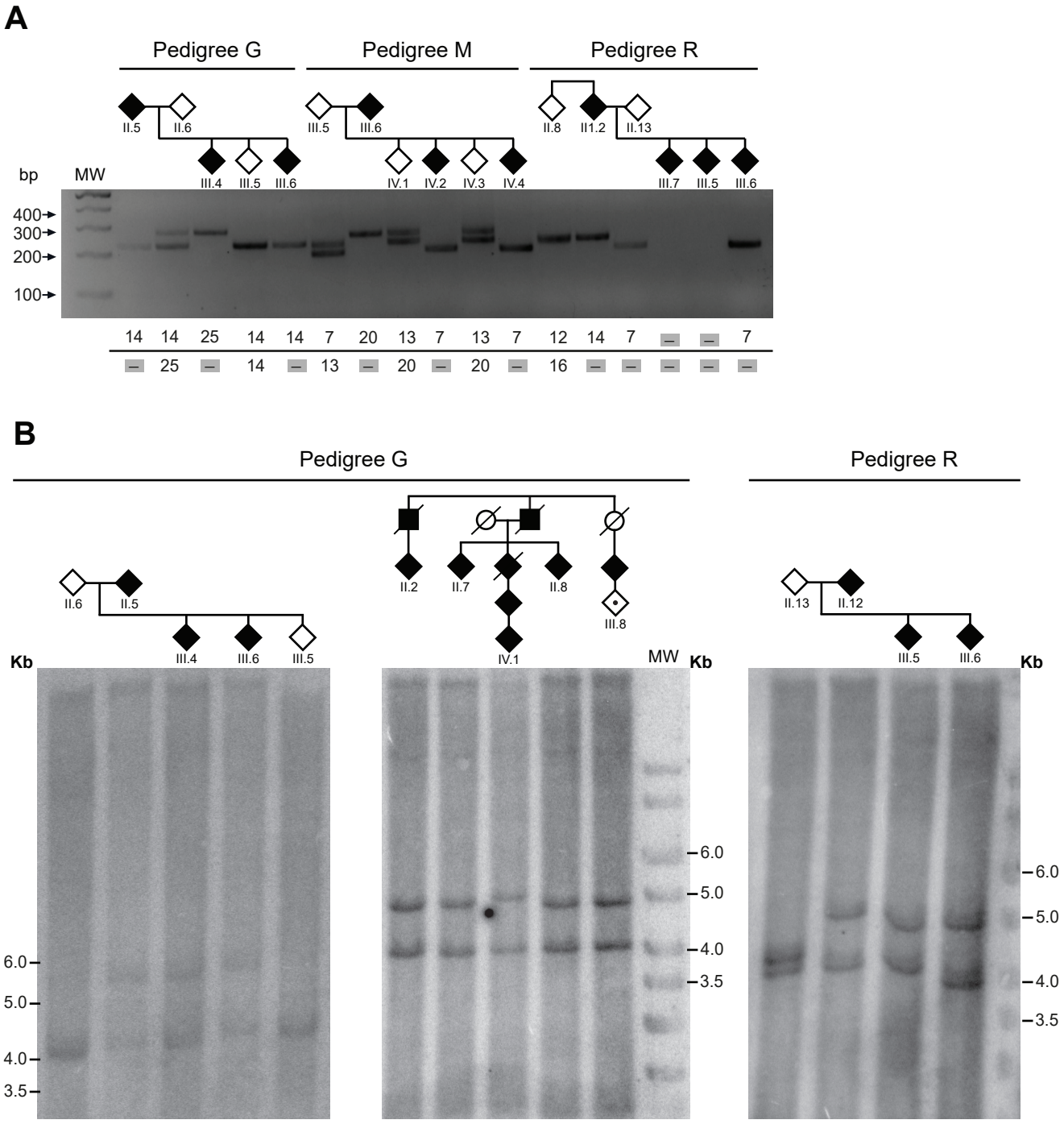


**Figure S1. PCR and Southern blot for (ATTTT)n in three large families.**

Related to **Figure 2**. (A) Standard PCR amplification shows no apparent $(ATTTT)_n$ transmission from all affected parent to affected offspring, $(ATTTT)_n$ size was assessed by Sanger sequencing and is indicated below each lane in the agarose gel image. (B) Southern blot analysis with a probe hybridizing near the pentanucleotide repeat ATTTT/AAAAT in *DAB1* showing a fragment of approximately 5.1 kb corresponding to ~200 pentanucleotide repeats that is transmitted from the affected parent to affected offspring. Individual ID number is according to individual ID number in each pedigree of Figure 1.

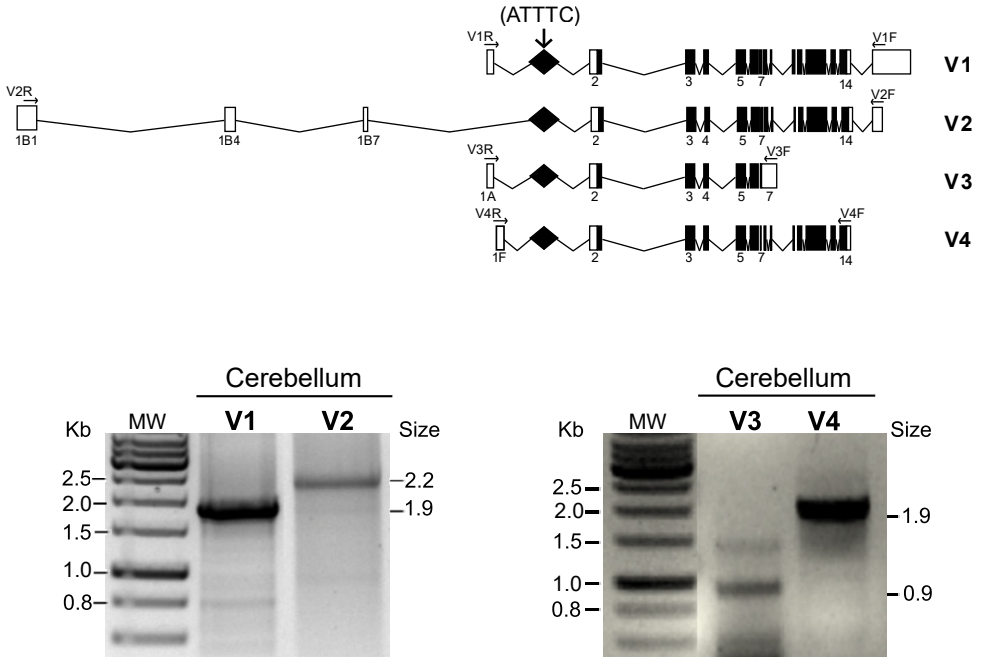


**Figure S2. Cerebellar Expression of *DAB1* transcripts with the (ATTTC)n Insertion Intron.**

Related to **Figure 6**. Bellow is the RT-PCR analysis of *DAB1* transcripts from human total cerebellum RNA (Clontech). PCR products correspond to amplification with specific primers, represented with arrows, in the schematic representation above of *DAB1* transcripts, for the first and last exons of variant 1 (V1), V2, V3 and V4.

Table S1. Primers used for PCR, Sequencing, Long range PCR, RP-PCR, Southern Blot, rt-PCR and T7 promoter cloning.

| Primer ID | Primer sequence (5'-3') |
|---|---|
| $(ATTTT)_n$ screening | |
| 24F | GAAGTGGTCCTCCCAAGTCA |
| 24R | ACACTTTGGGAGGCAGAGG |
| Sequencing primers | |
| 24F | GAAGTGGTCCTCCCAAGTCA |
| 24R4 | GAGACCAGCCTGGGCAAC |
| Long PCR | |
| ALU24F | ATTTGCCCTTTGCTGATTGA |
| ALU24R | TGAAACTGAGGCTCAAAATGA |
| RP-PCR | |
| 24R | [6FAM]ACACTTTGGGAGGCAGAGG |
| RP-TTCAT | TACGCATCCCAGTTTGAGACGTTCATTTCATTTCATTTCATTTCAT |
| FLAG | TACGCATCCCAGTTTGAGACG |
| *Alu*Jb $(ATTTT)_n$ and *Alu*Jb $(ATTTC)_n$ allele Cloning | |
| ALUEcoRIF | GTCAGTGAATTCAATTAATTTGCCCTTTGCTGATT |
| ALUNotIR | GTCAGTGCGGCCGCTGAAACTGAGGCTCAAAATGA |
| Southern Blot probe | |
| SB24F | CAGGAGGGAGGTGCTTCTG |
| SB24R | TCACACAAAATGGTACTCTGAAAAA |
| rt-PCR full length *DAB1* transcript variants | |
| V1F | GAGGAGGATGCTCTGGGCTA |
| V1R | TCCCAGACCTGCGCTATCTA |
| V2F | CCCAGGGAACAAAAGCGGA |
| V2R | TTGATCTGCGCGTACAGAGG |
| V3F | AGGAGGATGCTCTGGGCTAGG |
| V3R | CCAGGTTTTGCAGTCAGTGGA |
| V4F | GTTCGGCTCATAAAGCAGGC |
| V4R | CACTGGGCTCACCAAATGGA |
| T7 promoter cloning in pCDH-CMV-MCS-EF1-GFP-T2A-Puro | |
| T7A | CTAGGGATCCTAATACGACTCACTATAGGG |
| T7B | AATTCCCTATAGTGAGTCGTATTAGGATCC |

Table S3. Frequency in the Portuguese population of NGS variants

| Gene/Intergenic region | Variant | Position on Chr1 (hg19) | Frequency (1000g) | Homozygous variant (healthy individuals) | Heterozygous variant (affected individuals) | Chromosomes studied (N) | Frequency Portuguese population |
|---|---|---|---|---|---|---|---|
| Intergenic | rs565332393[a] | 56,090,535 | 0.0002 | T | C | 438 | 0.0046 |
| | rs762335464[a] | 56,223,397 | * | T | C | 222 | 0.0089 |
| | rs142969184 | 56,251,658 | 0.0018 | C | A | 101 | 0.0289 |
| | rs761751006[a] | 56,305,070 | * | A | G | 656 | 0.0030 |
| | ss2137493855[a,b] | 56,453,113 | * | A | G | 1311 | 0.0008 |
| | rs777060331[a] | 56,517,710 | * | T | C | 266 | 0.0075 |
| | ss2137493856[a,b] | 56,545,567 | * | A | G | 1330 | <0.002 |
| | rs138928773 | 56,753,932 | 0.0002 | T | C | 712 | 0.0028 |
| | rs528859858[a] | 56,810,200 | * | G | A | 430 | <0.002 |
| *PPAP2B* | rs537634498[a] | 56,964,113 | * | - | CCCAGC | 14 | 0.0714 |
| *C1orf168* | rs555296478 | 57,250,815 | 0.0014 | T | - | 752 | 0.0027 |
| *C8A* | rs572272180 | 57,367,559 | 0.0024 | TTG | - | 192 | 0.0052 |
| Intergenic | rs115293800 | 57,438,757 | 0.0039 | G | C | 162 | 0.0241 |
| | rs866411539[a] | 57,444,772 | * | G | T | 3052 | 0.0007 |
| *DAB1* | rs145962085 | 57,481,145 | 0.0038 | T | C | 252 | 0.0079 |
| | ss2137493861[a,b] | 57,491,966 | * | T | G | 2900 | 0.0007 |
| | ss2137493857[a,b] | 57,551,605 | * | G | A | 1534 | 0.0013 |
| | rs192485043 | 57,926,567 | 0.0002 | T | A | 661 | 0.0015 |
| | rs145097803 | 58,201,704 | 0.0032 | T | A | 434 | 0.0046 |
| | ss2137493862[a,b] | 58,215,160 | * | C | G | 3030 | 0.0007 |

[a]Variants not present in dbSNP135 and 1000 Genomes released in February 2012; [b]Submitted to NCBI, ss#; *Not available.

Table S4. Configuration of $(ATTTC)_n$ insertion alleles in the three additional families

| Family (Haplotype) | Affected individuals | $(ATTTT)_n$ allele | $(ATTTC)_n$ insertion allele |
|---|---|---|---|
| MS (B) | MS1 | $(ATTTT)_{19}$ | $(ATTTT)_{73}(ATTTC)_{46}(ATTTT)_{81}$ |
| | MS2 | $(ATTTT)_{19}$ | $(ATTTT)_{69}(ATTTC)_{52}(ATTTT)_{90}$ |
| | MS3 | $(ATTTT)_{19}$ | $(ATTTT)_{72}(ATTTC)_{46}(ATTTT)_{16}$[b] |
| | MS4 | $(ATTTT)_{12}$ | $(ATTTT)_{67}(ATTTC)_{61}(ATTTT)_{70}$[b] |
| C (C) | C1 | $(ATTTT)_{16}$ | $(ATTTT)_{81}(ATTTC)_{64}(ATTTT)_{64}$[b] |
| D (D) | D1 | $(ATTTT)_{16}$ | $(ATTTT)_{52}(ATTTC)_{71}(ATTTT)_{83}$[b] |

[b]3' ATTTTs not completely sequenced