The American Journal of Human Genetics, Volume *110*

Supplemental information

An intronic GAA repeat expansion in *FGF14* causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14

Haloom Rafehi, Justin Read, David J. Szmulewicz, Kayli C. Davies, Penny Snell, Liam G. Fearnley, Liam Scott, Mirja Thomsen, Greta Gillies, Kate Pope, Mark F. Bennett, Jacob E. Munro, Kathie J. Ngo, Luke Chen, Mathew J. Wallis, Ernest G. Butler, Kishore R. Kumar, Kathy HC. Wu, Susan E. Tomlinson, Stephen Tisch, Abhishek Malhotra, Matthew Lee-Archer, Egor Dolzhenko, Michael A. Eberle, Leslie J. Roberts, Brent L. Fogel, Norbert Brüggemann, Katja Lohmann, Martin B. Delatycki, Melanie Bahlo, and Paul J. Lockhart

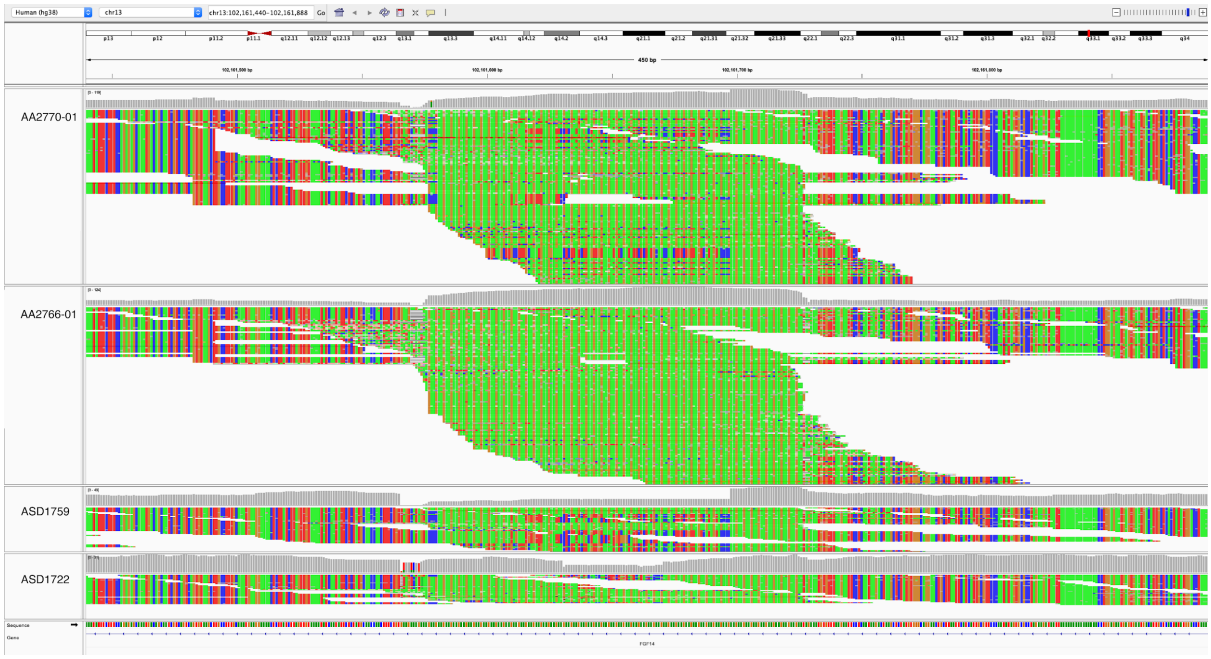


Figure S1: IGV snapshots of the $(GAA)_n$ locus in *FGF14*.

Visualization of the GS reads in IGV for individuals with an expanded (GAA) RE (top two images) compared to controls (bottom two images).

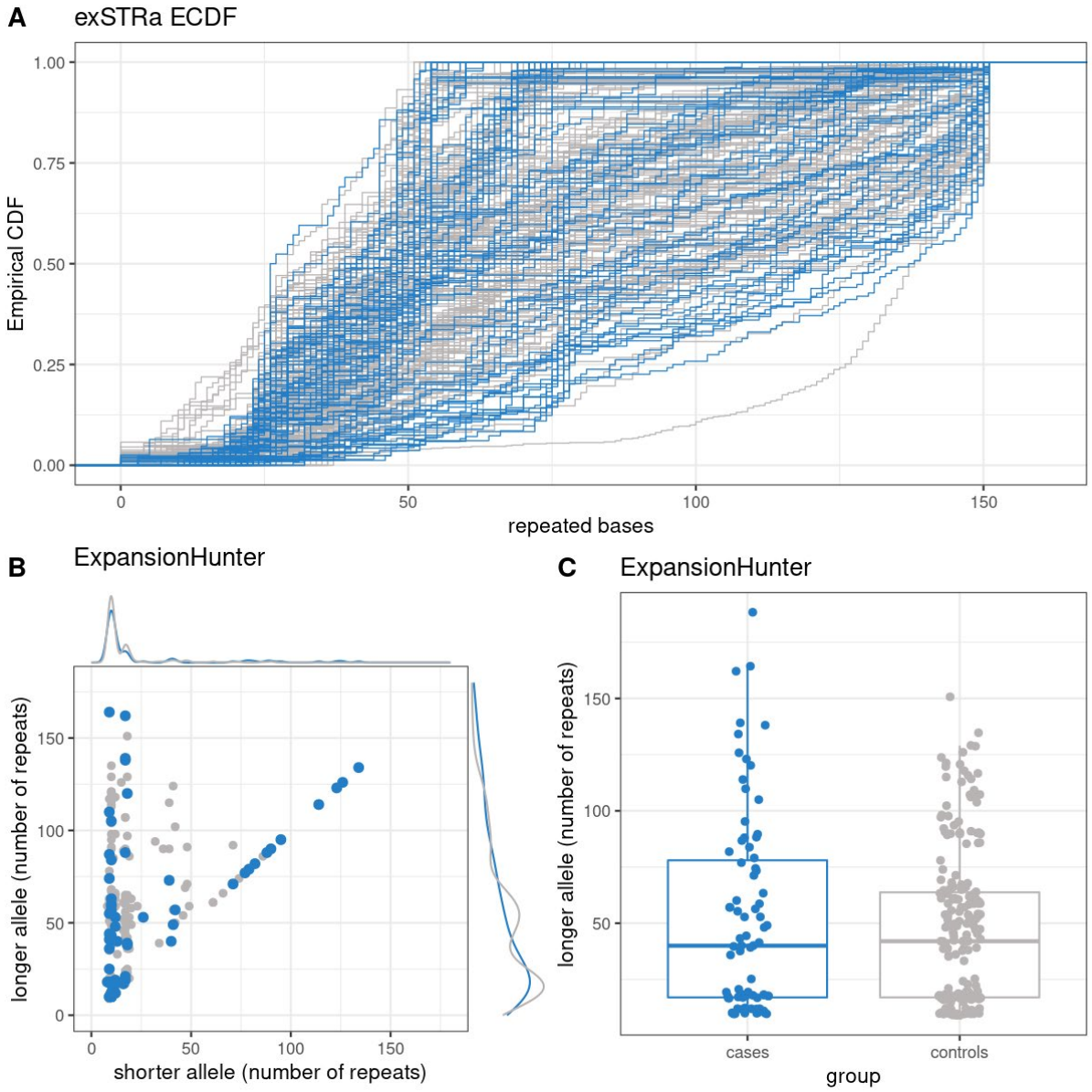


Figure S2: Plots of exSTRa and ExpansionHunter outputs for the *FGF14* (GAA) locus

The *FGF14* (GAA) STR was profiled in cases (blue) and controls (grey) using bioinformatics tools, showing inability of bioinformatic tools to accurately estimate the size of the *FGF14* STR, particularly for alleles with $>(GAA)_{100}$. (A) exSTRa ECDF plot fails to distinguish outliers, due to the high variability of and large size of the STR in the general population. (B) Size estimates of the *FGF14* (GAA) STR were determined using ExpansionHunter. Density plots are show the allele repeat distributions for cases (blue) and controls (grey) for the longer (y-axis) and shorter (x-axis) alleles. (C) Comparison of longer allele size determined by ExpansionHunter in cases versus controls.

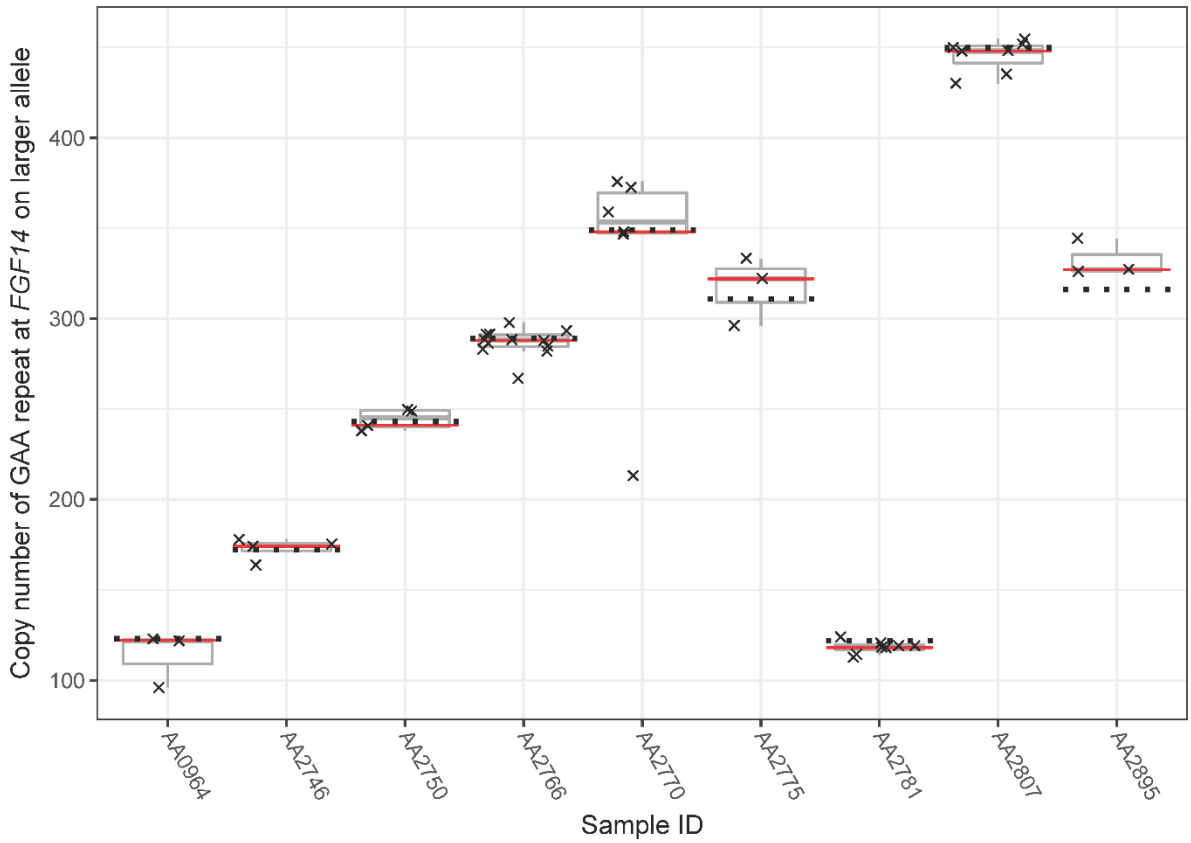
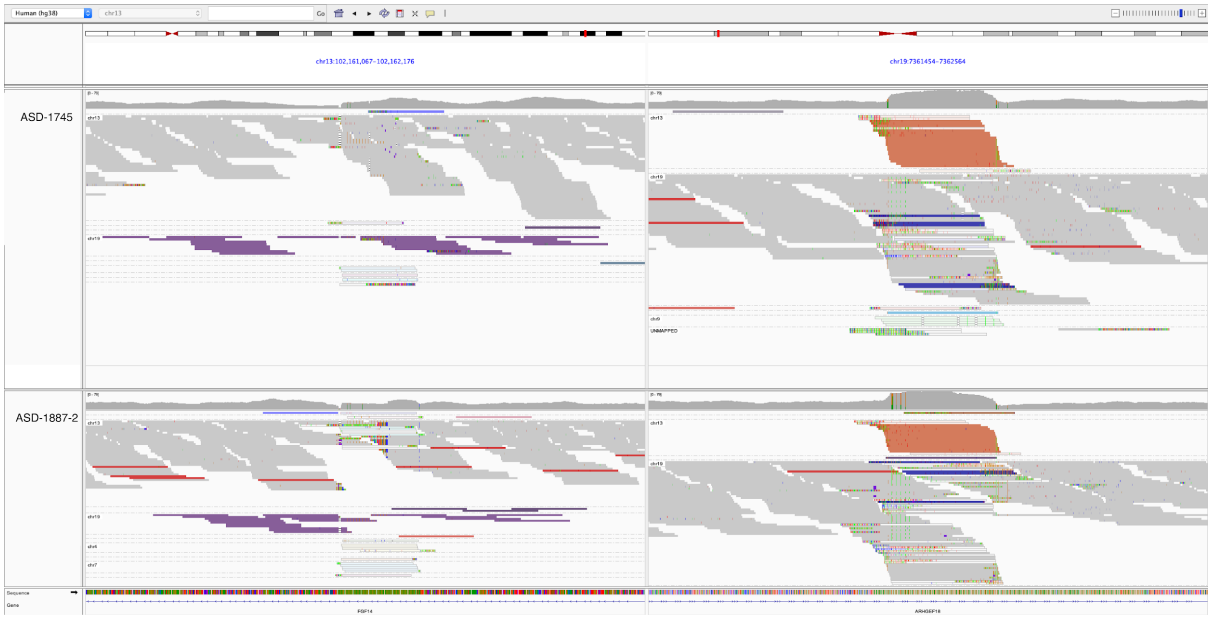


Figure S3: Nanopore repeat size estimates concur with PCR estimates. Eight probands were sequenced with Oxford Nanopore MinION adaptive sequencing targeting *FGF14*. The $(GAA)_n$ repeat was genotyped in the long-read sequencing data using tandem-genotypes. Repeat sizes from Nanopore sequencing are shown for each read (crosses) and as a distribution (box plot), along with the tandem-genotypes (red line) and PCR (dotted line) size estimates for each sample. All samples except AA0964, AA2764 and AA2781 encoded a pathogenic *FGF14* allele.



SF4: IGV snapshot of controls with unusual RP-PCR patterns

Visualization of the genome sequence reads in IGV for controls with an expanded GAAGGA RE at the *FGF14* $(GAA)_n$ STR (left panel). Reads coloured in purple uniquely map to the *FGF14* locus, however their read pairs map to a GAAGGA motif in another region of the genome (right panel, reads in orange).

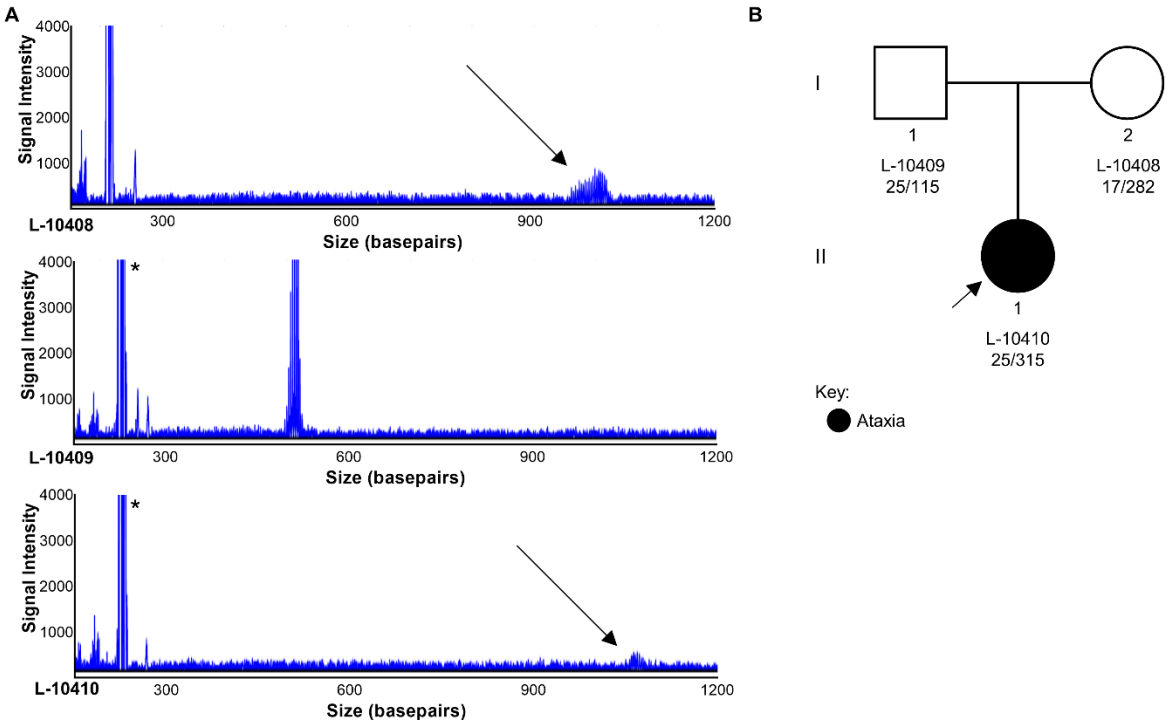


Figure S5: Intergenerational expansion of *FGF14* (GAA)n STR

LR-PCR electropherograms of *FGF14* STR by capillary array (A) for II-1 (L-10410) and unaffected parents (L-10408 and L-10409) (B). Allele sizing for L-10408 at this locus demonstrated that the paternal allele $(GAA)_{25}$ (indicated by *) was inherited unchanged, whereas the maternal allele increased from $(GAA)_{282}$ to $(GAA)_{315}$ in the affected daughter.

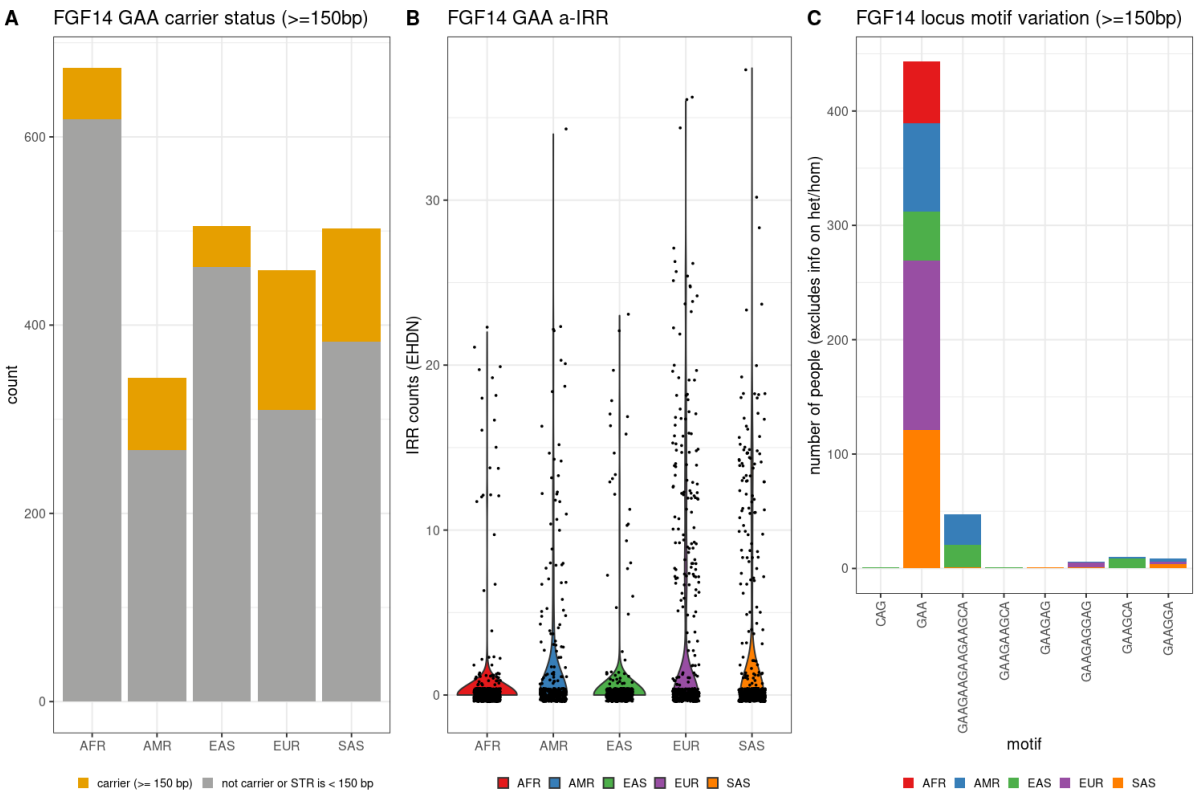


Figure S6: Profiling the *FGF14* (GAA) RE locus in the 1000 Genomes Project.
The *FGF14* (GAA) STR locus was profiled in 2483 unrelated individuals from the 1000 Genomes Project from African (AFR), ad mixed American (AMR), East Asian (EAS), European (EUR) and South Asian (SAS) populations. (A) Number of *FGF14* (GAA) STR carriers in the 1000 Genomes Project, by ethnicity. Carriers are defined as having an STR >= 150 bp (or 50 repeats). (B) Violin plot showing the distribution of the a-IRR for the *FGF14* (GAA) STR by ethnicity. Individuals without a reported a-IRR, either because they do not carry the STR, or they carry the STR with <150 bp (or 50 repeats), are coded as zero. (C) Bar chart showing the alternative motifs also detected at the *FGF14* (GAA) locus, split by ethnicity. Only motifs present at >= 150bp in at least one individual are reported.

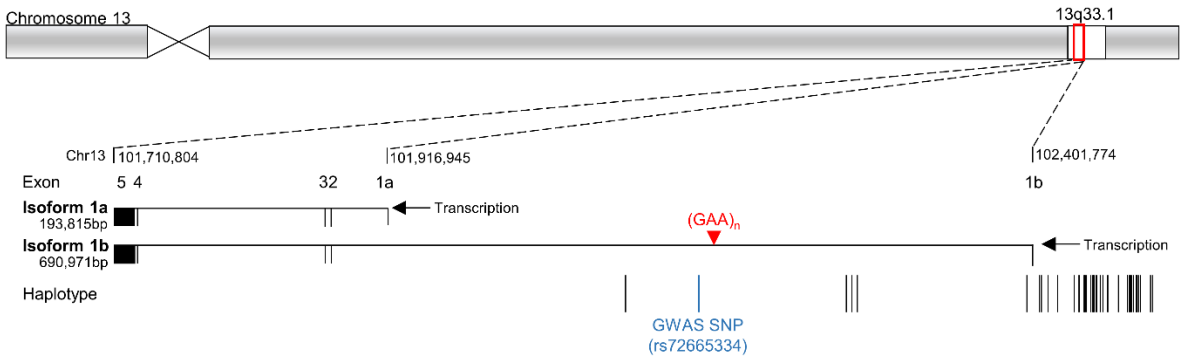


Figure S7. *FGF14* genomic region, haplotype and isoform structure

Schematic representation of the chromosomal location of *FGF14* to 13q33.1, illustrating the expression sites of two mRNA isoforms, 1a and 1b. Exon one of isoform 1b is located 5' of 1a, and the $(GAA)_n$ STR is located within the intervening intron. The remaining exons 2-5, are common to both isoforms. The SNP (rs72665334) identified by GWAS to be associated with down beat nystagmus is located ~10kb 3' of the $(GAA)_n$ STR. A common ancestral haplotype was determined for three Australian cases with $(GAA)_{>250}$ by comparing variant sharing around the *FGF14* STR (gnomAD AF<0.1) (specific haplotype details in Table S3).

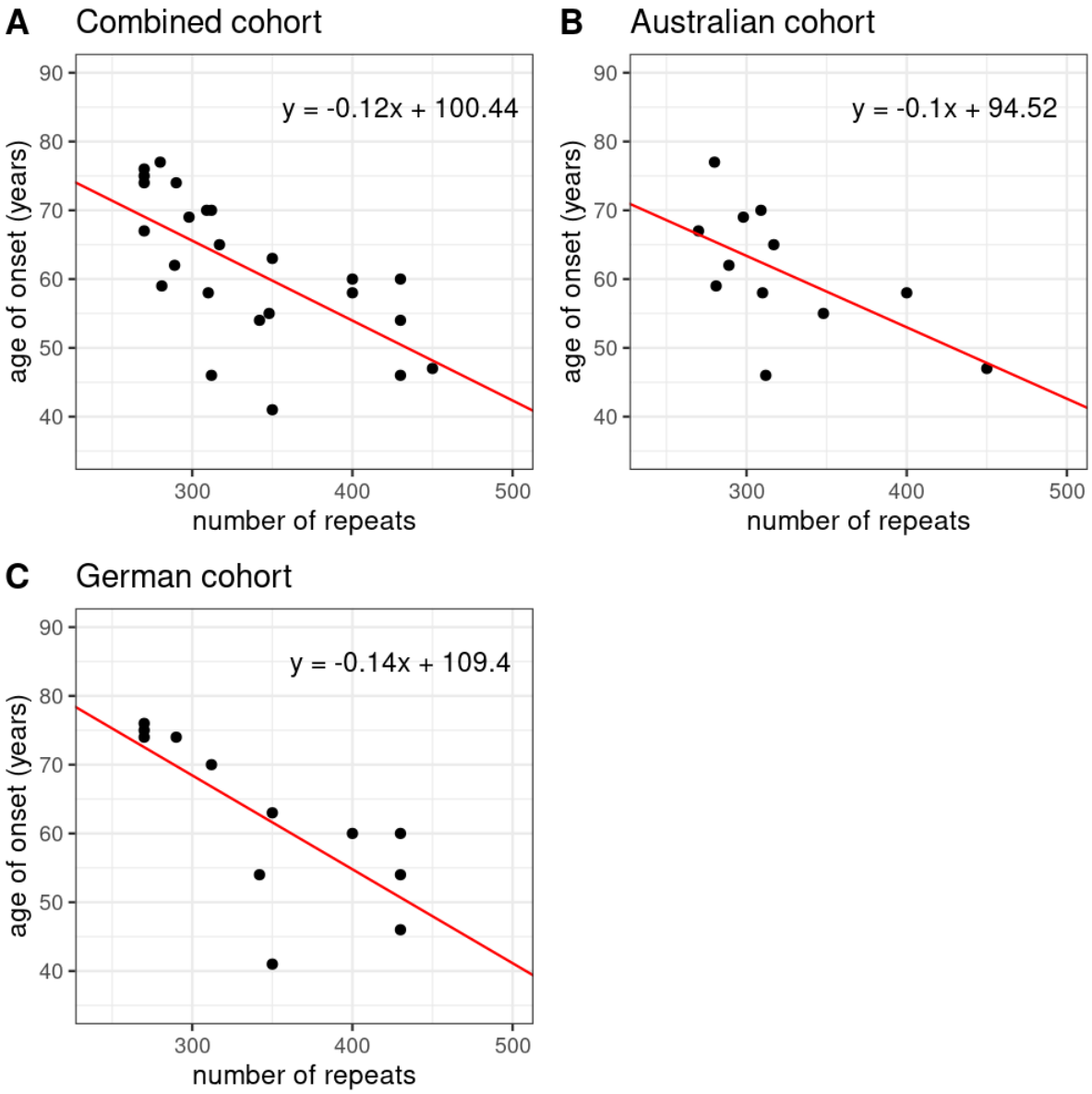


Figure S8: Linear regression comparing *FGF14* $(GAA)_n$ RE number of repeats and age at onset.

A linear regression was used to determine the relationship between ataxia age at onset and *FGF14* $(GAA)_n$ repeat length for the (A) combined (n=24), (B) Australian (n=12) and (C) German (n=12) cohorts. Repeat lengths were determined using LR-PCR. Individuals with ataxia of known age at onset with $(GAA)_{>250}$ were included in the analysis.

Table S1: Primer sequences for analysis of *FGF14*.

Primer name	Position (hg38)	Sequence
FGF14_RPP_ F1	chr13:102161468-102161490	5'-AGCAATCGTCAGTCAGTGTAAGC
FGF14_LRP_ R1	chr13:102161762-102161782	5'-CAGTTCCTGCCCACATAGAGC
FGF14_RPP_ F1_FAM	chr13:102161468-102161490	FAM-5'-AGCAATCGTCAGTCAGTGTAAGC
FGF14_RPP_AAG_RE_R1	NA	5'-CAGGAAACAGCTATGACC CTTCTTCTTCTTCTTCTTCTT
RPP_M13R	NA	5'-CAGGAAACAGCTATGACC

The gene reference sequences utilized for *FGF14* were NC_000013.11, NM_004115.4 (Isoform 1a) and NM_175929.3 (Isoform 1b).

Table S2. *FGF14* allele sizing for individuals with $(GAA)_{>250}$ in Australian and German cohorts

Australian Cohort						
Affected Individuals			Controls			
Alias	Allele		Alias	Age (years)	Allele	
	1	2			1	2
AA2807	170	450	ASD1745	48	65	332*
AA2831	74	400	C0063	61	12	300
AA2903	19	400	ASD1808	43	12	296*
AA2770	13	349	ASD1887	38	22	282*
AA2895	58	316	C0090	46	12	270
AA0441	19	315				
AA2809	12	313				
AA2775	57	311				
AA2845	19	296				
AA2766	185	289				
AA2933	20	284				
AA2926	12	268				
AA2908	14	256				

German Validation Cohort						
Affected Individuals			Controls			
Alias	Allele		Alias	Age (years)	Allele	
	1	2			1	2
L-18362	28	460	L-3657	52	49	332
L-14575	29	460	L-3656	56	315	330
L-18384	19	430	L-3501	74	19	329
L-17672	20	430	L-3344	54	54	328
L-17665	29	430	D11	49	100	325
L-20363	54	400	L-3479	51	52	325
L-15166	19	350	L-3013	65	19	325
L-15764	87	350	A10	87	19	290
L-14630	19	342	C3	49	27	282
L-10410	25	315	A5	42	19	257
L-15891	54	312				
L-15754	19	290				
L-15629	19	270				
L-15739	59	270				
L-15713	19	270				

*These samples have a $(GAAGGA)_n$ motif and are considered non-pathogenic.

Table S3. Summary of *FGF14* core haplotype shared by three Australian individuals with $(GAA)_{>250}$.

chr	POS	REF	ALT	SYMBOL	AF	Rs name	AA2766	AA2770	AA2831
chr13	102095119	C	T	FGF14	0.008	rs79181669	0/1	0/1	0/1
chr13	**102150076**	**C**	**T**	**FGF14**	**0.0218**	**rs72665334**	**0/1**	**0/1**	**0/1**
chr13	**102161575**	**$(GAA)_{50}$**	**$(GAA)_{>250}$**	**FGF14**	**-**	**-**	**$(GAA)_{289}$**	**$(GAA)_{348}$**	**$(GAA)_{>400}$**
chr13	102261142	G	A	FGF14	0.0154	rs117891033	0/1	0/1	0/1
chr13	102264941	T	C	FGF14	0.0297	rs72647446	0/1	0/1	0/1
chr13	102269434	C	T	FGF14	0.0291	rs72647448	0/1	0/1	0/1
chr13	102396542	G	C	FGF14	0.0771	rs9557860	0/1	0/1	0/1
chr13	102405672	A	G	FGF14	0.0891	rs72649411	0/1	0/1	0/1
chr13	102407597	C	T		0.0944	rs72649412	0/1	0/1	0/1
chr13	102412712	T	G		0.0903	rs17690152	0/1	0/1	0/1
chr13	102419917	G	T		0.0879	rs58913166	0/1	0/1	0/1
chr13	102432196	G	A		0.0914	rs61593134	0/1	0/1	0/1
chr13	102435740	A	G		0.091	rs2182843	0/1	0/1	0/1
chr13	102439745	C	T		0.0228	rs4772471	0/1	0/1	0/1
chr13	102440446	T	C		0.0795	rs114317848	0/1	0/1	0/1
chr13	102440534	G	A		0.099	rs66678179	0/1	0/1	0/1
chr13	102441121	T	C		0.0795	rs72649438	0/1	0/1	0/1
chr13	102444527	G	A		0.0797	rs72649442	0/1	0/1	0/1
chr13	102446132	T	C		0.0801	rs57911285	0/1	0/1	0/1
chr13	102446353	A	T		0.0781	rs57461252	0|1	0|1	0|1
chr13	102446354	A	C		0.0795	rs60040500	0|1	0|1	0|1
chr13	102447030	C	T		0.0863	rs116926265	0/1	0/1	0/1
chr13	102447183	A	G		0.0801	rs117258042	0/1	0/1	0/1
chr13	102448603	G	C		0.0801	rs1927360	0|1	0/1	0/1
chr13	102448793	A	T		0.0801	rs1927361	0/1	0/1	0/1
chr13	102451762	C	T		0.0799	rs55665569	0/1	0/1	0/1
chr13	102453431	A	G		0.0813	rs12585989	0/1	0/1	0/1
chr13	102457322	T	G		0.0801	rs55686828	0/1	0/1	0/1
chr13	102465342	T	C		0.0825	rs55638789	0/1	0/1	0/1
chr13	102472363	G	A		0.0731	rs72649478	0|1	0|1	0|1
chr13	102473510	G	T		0.0996	rs72649480	0/1	0/1	0/1
chr13	102474084	A	G		0.091	rs114125113	0/1	0/1	0/1
chr13	102474982	A	T		0.0968	rs77919312	0|1	0|1	0|1
chr13	102476968	C	T		0.0729	rs56164689	0/1	0/1	0/1
chr13	102478700	C	T		0.0713	rs72649487	0/1	0/1	0/1
chr13	102478941	T	C		0.0783	rs72649488	0/1	0/1	0/1
chr13	102479117	C	T		0.0783	rs55850322	0/1	0/1	0/1
chr13	102480213	T	C		0.0713	rs56287885	0/1	0/1	0/1
chr13	102481686	A	AG		0.0723	rs55728938	0/1	0/1	0/1
chr13	102489485	C	T		0.0951	rs17634018	0/1	0/1	0/1
chr13	102491020	T	C		0.0649	rs17634060	0/1	0/1	0/1

Supplementary Materials and Methods

RNA and protein analyses

Primary fibroblasts were established from three affected individuals and four unrelated sex and age matched controls using standard techniques. Cells were cultured as previously described[1] with RNA extracted (RNeasy Mini Kit, #74104, QIAGEN) and cDNA generated (QuantiTect Reverse Transcription Kit, #205313, QIAGEN) according to the manufacturer's instructions. Gene expression was analyzed using TaqMan Gene Expression Assays designed to identify all *FGF14* mRNA isoforms (Hs00738588, Thermo Fisher Scientific) or specifically the brain isoform 1b (Hs02888324, Thermo Fisher Scientific) according to the manufacturer's protocol. Protein was isolated and analyzed by Western blot as previously described[2] utilizing an antibody directed against FGF14 (UC Davis/NIH NeuroMab, #75-096).

Supplementary References

1. Wilson, G.R., Sunley, J., Smith, K.R., Pope, K., Bromhead, C.J., Fitzpatrick, E., Di Rocco, M., van Steensel, M., Coman, D.J., Leventer, R.J., et al. (2013). Mutations in SH3PXD2B cause Borrone dermato-cardio-skeletal syndrome. European journal of human genetics : EJHG. 10.1038/ejhg.2013.229.
2. Barbier, M., Bahlo, M., Pennisi, A., Jacoupy, M., Tankard, R.M., Ewenczyk, C., Davies, K.C., Lino-Coulon, P., Colace, C., Rafehi, H., et al. (2022). Heterozygous PNPT1 Variants Cause Spinocerebellar Ataxia Type 25. Annals of neurology *92*, 122-137. 10.1002/ana.26366.