SUPPLEMENTARY MATERIAL

Variants of the *elongator protein 3* (*ELP3*) gene are associated with motor neuron degeneration

Claire L Simpson[1], Robin Lemmens[2,3], Katarzyna Miskiewicz[4,5], Wendy J Broom[6],Valerie K Hansen[1], Paul WJ van Vught[7], John E Landers[6], Peter Sapp[6,8], Ludo Van Den Bosch[2,3], Joanne Knight[9], Benjamin M Neale[9], Martin R Turner[1], Jan H Veldink[7], Roel A Ophoff[10,11], Vineeta B Tripathi[1], Ana Beleza[1], Meera N Shah[1], Petroula Proitsi[12], Annelies Van Hoecke[2,3], Peter Carmeliet[3,13], H Robert Horvitz[8], P Nigel Leigh[1], Christopher E Shaw[1], Leonard H van den Berg[7], Pak C Sham[14], John F Powell[12], Patrik Verstreken[4,5], Robert H Brown Jr[6], Wim Robberecht[2,3], Ammar Al-Chalabi[1*].

1. MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, Department of Neurology, London SE5 8AF, UK

2. Service of Neurology (University Hospital Leuven) and Laboratory for Neurobiology, Section of Experimental Neurology, University of Leuven, Leuven, B-3000, Belgium

3. Vesalius Research Center, Flanders Institute for Biotechnology (VIB), Leuven, B-3000, Belgium

4. VIB, Department of Molecular and Developmental Genetics, Leuven, B-3000, Belgium

5. KULeuven, Center for Human Genetics, Laboratory of Neuronal Communication, Leuven, B-3000, Belgium

6. Cecil B Day Laboratory for Neuromuscular Research, Massachusetts General Hospital East, Charlestown, MA, USA

7. Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Netherlands

8. Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

9. MRC Social, Genetic and Developmental Psychiatry Centre, King's College London, Institute of Psychiatry, London SE5 8AF, UK

10. Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, Netherlands

11. Neuropsychiatric Institute, University of California, Los Angeles, USA

12. MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, Department of Neuroscience, London SE5 8AF, UK

13. Vesalius Research Center, KU Leuven, Leuven, B-3000, Belgium

14. Department of Psychiatry and Genome Centre, University of Hong Kong, Hong Kong

*Address for correspondence:

MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry P 043, London, SE5 8AF, UK

Email: ammar@iop.kcl.ac.uk

Telephone: +44 20 7848 5172, Fax: +44 20 7848 5190

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

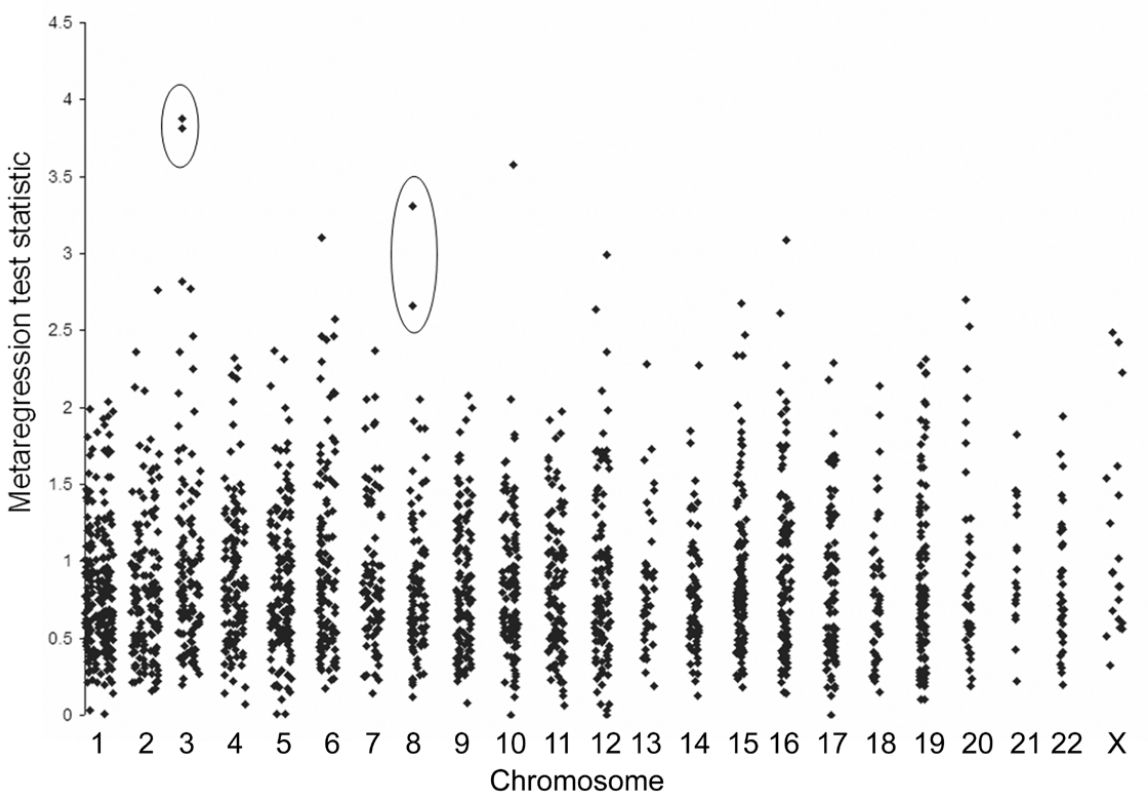

Figure S1. Results of analysis of microsatellite markers for association in UK population.

The four markers followed up by individual genotyping are circled. Both pairs were about 1Mb apart with at least one of the pair in the top four markers.

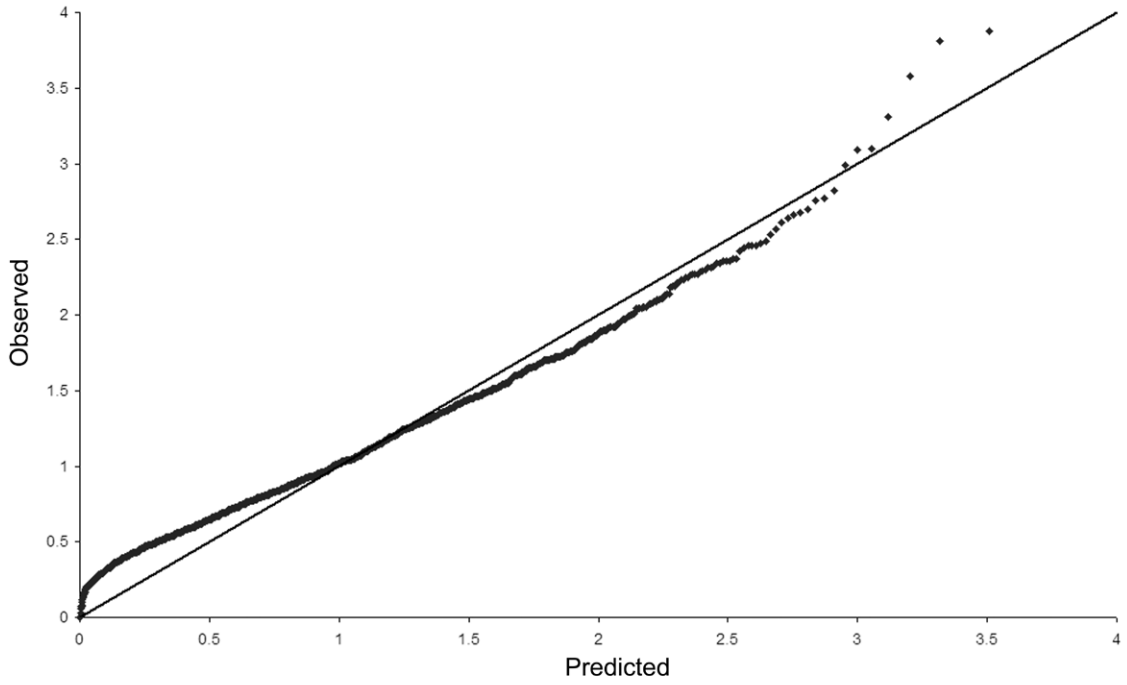

Figure S2. A Q-Q plot of the metaregression results for 1,884 microsatellite markers. Because allele frequencies were estimated from DNA pools, the test statistic has been deflated by variance estimates of the stutter artifact, differential amplification artifact, and sampling error for each marker.

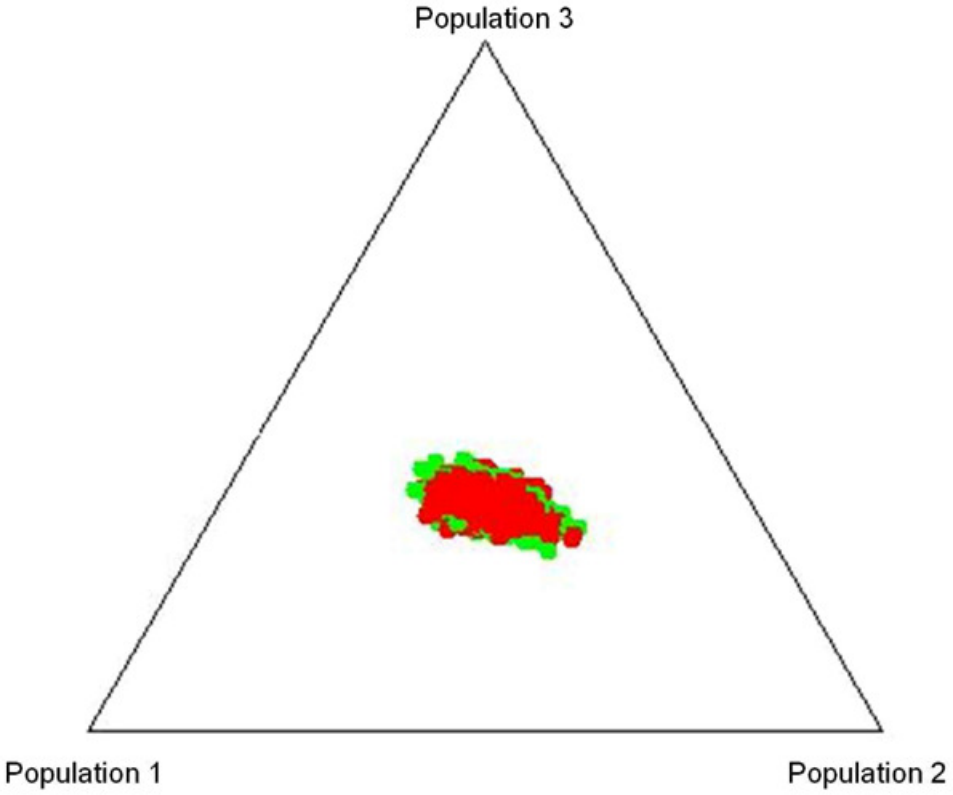

Figure S3. A triangle plot of the genetic distance for each individual from each of three putative ancestral populations (*K* = 3) estimated with the program Structure. Cases are shown in red and controls in green. The central grouping of plotted points suggests that the three populations are genetically similar. (Also see Figure S4).

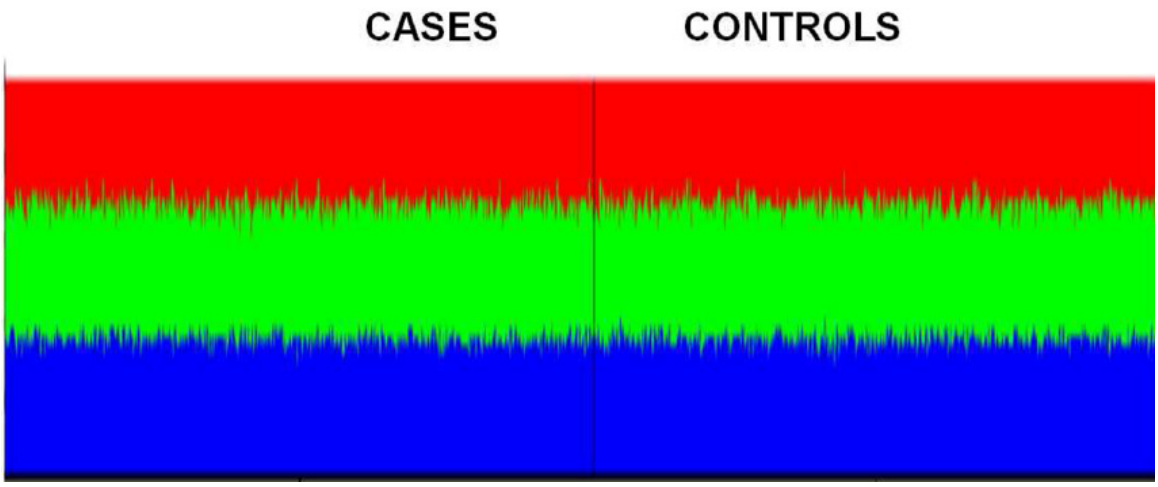

Figure S4. A bar plot of the inferred ancestry (Q) for each individual assuming three ancestral populations (*K* = 3) estimated with the program Structure.

The populations were genetically similar based on the χ^2 sum statistic for the 99 genomic control SNPs ($P = 0.57$) or analysis with the program Structure (43). In addition, assuming a single population, Wright's coefficient(44), F_{ST}, was well below the range for intracontinental populations at 0.0022. In this figure, each individual is plotted as a single vertical line partitioned into K colored segments representing that individual's estimated membership fraction in each of the K inferred clusters. The left of the plot shows cases and the right controls. All individuals are admixed with roughly a third (1/*K*) inferred ancestry from each putative population. This suggests there is no strong population structure. (Also see Figure S3).

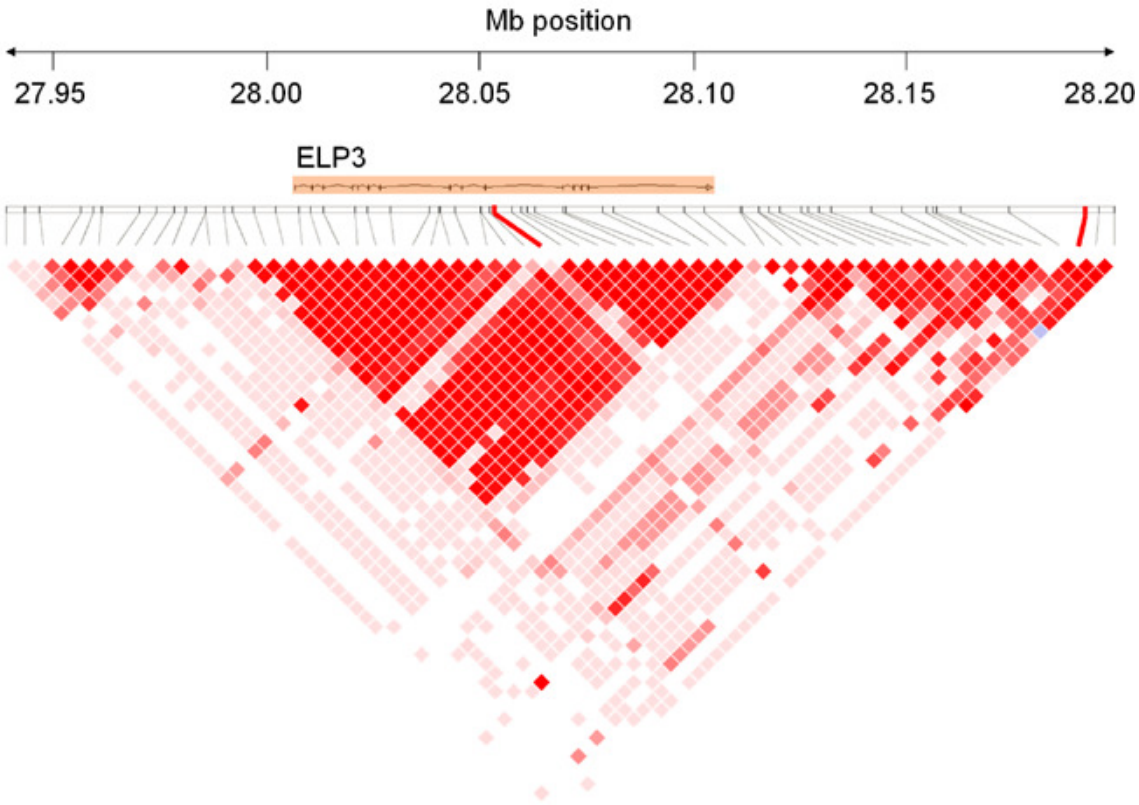

Figure S5. A linkage disequilibrium map of the *ELP3* region.

The genetic map in the region of *ELP3* has been plotted with the relative positions of *ELP3* shown in the center and genotyped markers shown as vertical lines descending to the triangle plot below. The position of marker D8S1820 is shown in red at 28.05 Mb. The shaded triangular region shows the D' values for LD for any pair of markers, with red being high values of D', shaded to white low values. For D8S1820 the commonest allele, 5, is shown as only one allele can be plotted on such graphs. The same map with allele 6 is provided in Figure S6 and a scatterplot to show all alleles is provided in Figure 1. Alleles of D8S1820 and rs12682496 (shown in red near 28.20 Mb) form a haplotype associated with a reduced risk of ALS.

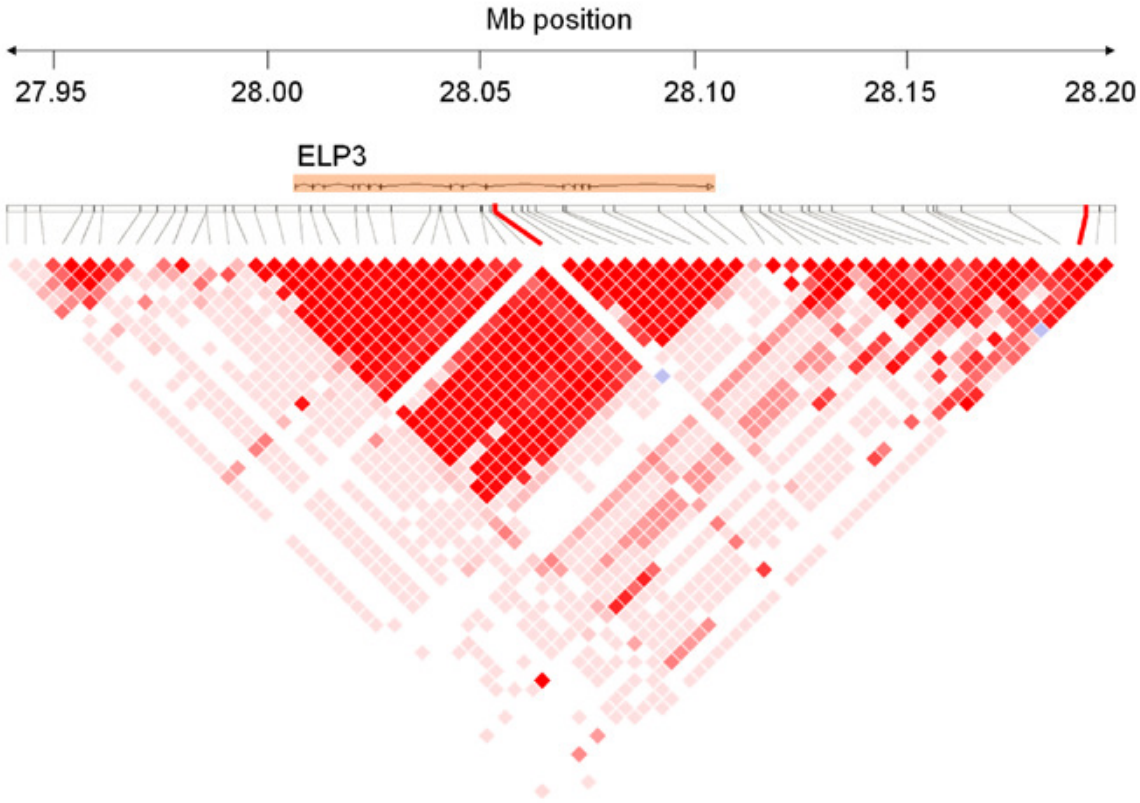

Figure S6. A linkage disequilibrium map of the *ELP3* region.

This figure is the same as Figure S5 except that allele 6 of D8S1820 is shown.

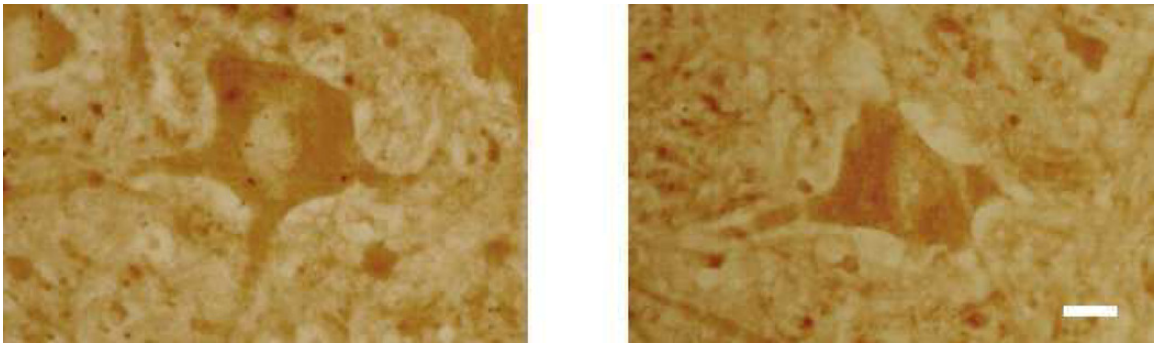

Figure S7. ELP3 expression in motor neurons of human spinal cord.

Paraffin sections of control human lumbar spinal cord stained for ELP3. The motor neuron in the center of each image stains brown confirming that ELP3 was present in motor neurons in the lumbar spinal cord. Scale bar 20 μm.

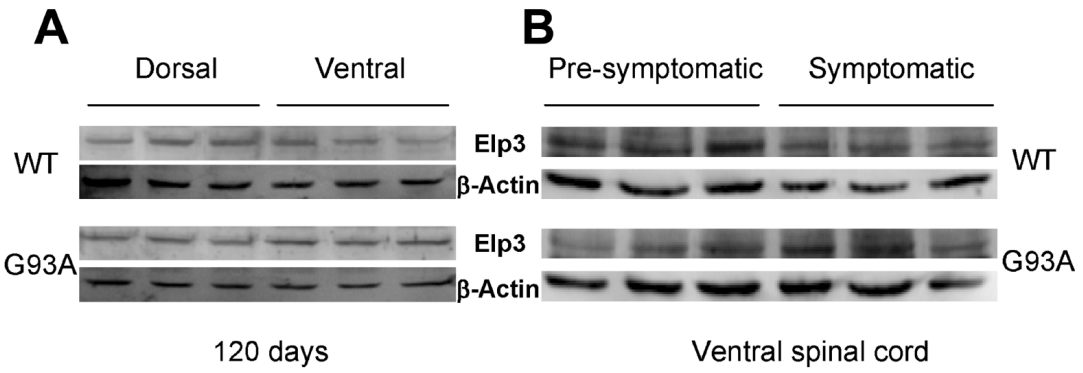

Figure S8. ELP3 expression in spinal cords of transgenic *SOD1*^{G93A} and *SOD1*^{WT} mice.

(A) ELP3 was present in ventral as well as dorsal spinal cord of *SOD1*^{G93A} and *SOD1*^{WT} mice at 120 days. There was no difference between dorsal and ventral spinal cords or between WT and G93A mice.

(B) In the ventral horn of G93A mice, ELP3 expression did not change after onset of disease: 60-80 days (pre-symptomatic) versus 120 days (symptomatic). Expression in WT mice at the equivalent ages is given for comparison (WT mice remain asymptomatic but the same labels have been used for ease of reference).

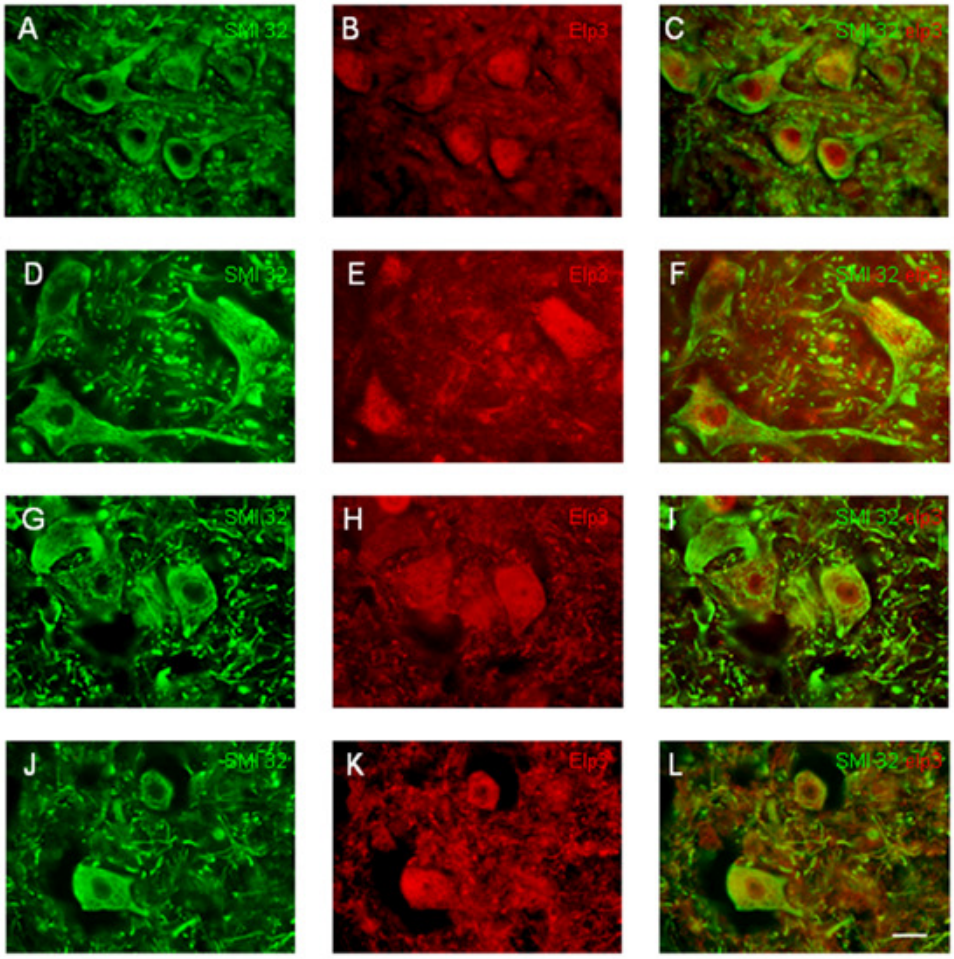

Figure S9. *ELP3* expression in motor neurons of spinal cords of transgenic *SOD1*^{G93A} and *SOD1*^{WT} mice demonstrated by immunostaining.

Fresh frozen spinal cord of SOD^{WT} mice (A)-(F) and SOD^{G93A} (G)-(L) were sectioned at presymptomatic (WT: (A)-(C) and G93A: (G)-(I)) and symptomatic (WT: (D)-(F) and G93A: (J)-(L)) ages. To identify motor neurons, sections were double stained with SMI32 antibody, which labels nonphosphorylated epitopes on medium and heavy subunits of neurofilament proteins (shown in green, (A), (D), (G) and (J)). Immunostaining for ELP3 showed ELP3 was present in motor neurons at different ages and in both transgenic animal stains (shown in red, (B), (E), (H) and (K)). Overlays are shown in panels (C), (F), (I) and (L). Scale bar 20 µm.

Table S1. Demographic characteristics of the samples studied.

Table showing the mean, standard deviation and range of age of onset of ALS in years as defined by first symptoms of weakness. For controls, age at sample donation was used.

	UK		***USA***		***Belgium***	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
Mean age of onset	56.4	56.4	53.78	56.6	57.3	40.2
Range	26.4-80.9	26.1-80.0	18.0-88.0	23.7-95.2	18.0-84.0	0.0-97.0
Standard deviation	12.5	12.5	13.5	14.1	12.5	18.4
Percentage male	59.7	59.7	64.6	35.9	62.4	44.2
Number of individuals	287	290	304	202	190	210

Table S2. Genetic association results for prioritized microsatellite markers.

Four markers were selected for initial follow-up. Although there were suggestive results for three of the markers in the overall analysis, only D8S1820 showed replicated association in all three populations.

	Rank	**Conventional genotyping, permuted *P*-values[a]**			**Total**
	Pools	**Individuals**	**Individuals**	**Individuals**	**Stratified test[b]**
Marker	**UK**	**UK**	**USA**	**Belgium**	
D3S1298	1	0.004	0.20	0.009	5×10^{-4}
D3S1260	2	0.003	0.83	0.61	0.048
D8S1048	4	2.20×10^{-5}	0.46	0.04	5.02×10^{-5}
D8S1820	12	9.00×10^{-4}	0.0023	2.60×10^{-6}	1.96×10^{-9}

[a]P-values were calculated by up to 10,000,000 permutations using all 15 alleles in the program CLUMP. [b]For the stratified test, P-values were combined with Fisher's method.

Table S3. Counts and frequencies for ELP3 alleles in cases and controls.

There are eight common and seven rare alleles of the D8S1820 microsatellite in intron 10 of *ELP3*. Allele 13 was assumed to exist based on repeat structure, although no individuals carrying it were identified. The classification into protection-associated or risk-associated alleles was carried out by CLUMP.

Allele	**Size bp**	**Count cases**	**Count controls**	**Frequency cases**	**Frequency controls**	**Risk- or protection-associated**
1	90	0	2	0	0.001	protection-associated
2	92	3	0	0.002	0	risk-associated
3	94	2	1	0.001	0.001	risk-associated
4	96	26	18	0.017	0.013	risk-associated
5	98	458	387	0.294	0.276	risk-associated
6	100	42	80	0.027	0.057	protection-associated
7	102	11	10	0.007	0.007	risk-associated
8	104	391	358	0.25	0.255	risk-associated
9	106	527	426	0.337	0.303	risk-associated
10	108	51	75	0.033	0.053	protection-associated
11	110	48	38	0.031	0.027	risk-associated
12	112	3	3	0.002	0.002	risk-associated
13	114	0	0	0	0	
14	116	0	2	0	0.001	protection-associated
15	118	0	4	0	0.003	protection-associated
TOTAL		1562	1404			

Table S4. Markers typed in and around the *ELP3* gene.

Table showing results of tests for association between SNP markers around *ELP3* and ALS. These results are consistent with those from genome-wide association studies and are probably a result of the poor LD between the most strongly associated alleles of D8S1820 and nearby SNPs.

Number	**Marker name**	**Base pair position**	χ^2 ***P*-value**[a]
1	rs7815173	27939020	0.6158
2	rs4732803	27943494	0.4817
3	rs1377336	27946777	0.2671
4	rs11780888	27956629	0.3842
5	rs11136028	27959504	0.7257
6	rs1000658	27961518	0.195
7	rs12681850	27970396	0.411
8	rs12549429	27974287	0.1133
9	rs2045029	27978493	0.9692
10	rs4732620	27981123	0.5852
11	rs13271784	27985846	0.4497
12	rs9644136	27985898	0.2807
13	rs1901744	27990118	0.9521
14	rs4732815	27992167	0.2648
15	rs17487788	27997263	0.4932
16	rs1377338	28002292	0.3665

17	rs2167768	28007635	0.8796
18	rs1563055	28010384	0.3893
19	rs2290369	28013475	0.909
20	rs6996985	28020775	0.852
21	rs3213997	28023775	0.6597
22	rs4732823	28029134	0.9107
23	rs11136032	28038235	0.9763
24	rs1530929	28040396	0.8162
25	rs1000275	28040848	0.2805
26	rs7812540	28044231	0.8529
27	rs4732828	28050160	0.7604
28	rs4732628	28050420	0.7276
29	rs3757895	28052285	0.6284
30	D8S1820	28053522	1.96×10^{-9}
31	rs10104739	28057599	0.9936
32	rs7003418	28059693	0.8559
33	rs10095097	28061229	0.8826
34	rs10216910	28062983	0.6333
35[b]	*rs9644138*	*28066473*	*0.8965*
36	rs4732832	28069498	0.9926
37	rs2322899	28070295	0.9886
38	rs2015443	28078845	0.7387
39	rs351761	28081444	0.9312

40	rs1350801	28091897	0.915
41	rs6558044	28098029	0.8461
42	rs7833735	28102730	0.9763
43	rs13268953	28111155	0.02857
44	rs4732631	28111472	0.4645
45	rs4054817	28115460	0.9644
46	rs4319045	28115564	0.7425
47	rs13253111	28117893	0.944
48	rs12675992	28120336	0.274
49	rs12544946	28125347	0.6575
50	rs6986290	28127038	0.4625
51	rs17411951	28129563	0.5138
52	rs6985192	28132341	0.2856
53	rs10094977	28141727	0.2256
54	rs6558049	28148960	0.4286
55	rs6987991	28154703	0.8748
56	rs13277389	28156153	0.5208
57	rs13254276	28157158	0.8974
58	rs2614114	28162703	0.4136
59	rs2100814	28174049	0.7637
60	rs12682496	28192028	0.3375
61	rs1461223	28195157	0.8744
62	rs2614048	28198702	0.3999

[a]*P*-values are for a Mantel-Haenzsel χ^2 test for association with ALS except in the case of D8S1820 where the Fisher method for combining p-values was used to combine permutation-based *P*-values derived from the program CLUMP. [b]Alleles of marker 35 were not in Hardy-Weinberg equilibrium.

Table S5. Elements of the RNA processing pathway implicated in neurodegeneration.

Table showing some of the proteins linked, associated or pathologically implicated in neurodegeneration. Proteins from the RNA processing pathway appear to be particularly important for motor neurons.

RNA process	**Protein**	**MIM**	**Complex/function**	[b]**Human disease**	**Reference**
Initiation					
? pre-mRNA	SETX	608465	DNA/RNA helicase	ALS4	(24)
? pre-mRNA	IGHMBP2	600502	DNA/RNA helicase	SMARD/HMN6	(45)
pre-mRNA[a]	*ELP3*		*Histone acetylation*	*ALS*	
pre-mRNA	ATXN7	607640	GCN5/TFTC/STAGA	SCA7	(27)
Elongation					
pre-mRNA	IKAP	603722	RNA polymerase II	FD	(26)
pre-mRNA	ELP2		RNA polymerase II		
pre-mRNA[a]	*ELP3*		*RNA polymerase II*	*ALS*	
pre-mRNA	ELP4, 5 and 6	606985	RNA polymerase II	Aniridia	
Termination					
Splicing					
pre-mRNA	UNRIP	605986	Spliceosome		
mRNA	HSP73	600816	Protection from RNase		
pre-mRNA	SMN	600354	Spliceosome	SMA	(46)
pre-mRNA	Gemin2-8	602595	Spliceosome		
pre-mRNA	TDP43	605078		ALS	(2, 47)

Translation					
tRNA[a]	*ELP3*		*tRNA synthesis*	*ALS*	
tRNA	GARS	600287	tRNA synthesis	CMT2D, HMN5	(48)
RNA	ANG	105850	RNase	ALS	(25)

[a]The three known functions of ELP3 are shown in italics. [b]SMARD, SMA with respiratory distress; CMT, Charcot-Marie-Tooth disease; HMN, Hereditary motor neuronopathy; SCA, spinocerebellar ataxia FD, familial dysautonomia; SMA, spinal muscular atrophy.