Supplementary material for Steinberg KM, et al. Whole genome analyses reveal no pathogenetic single nucleotide or structural differences between monozygotic twins discordant for amyotrophic lateral sclerosis. Amyotroph Lateral Scler and Frontotemporal Degeneration, 2015; DOI: 10.3109/21678421.2015.1040029.

Whole genome sequencing methods

The yield and integrity of native genomic DNA was verified by a PicoGreen assay for quantitation (Invitrogen) and a 0.8% Agarose gel for a qualitative QC. Samples were also run on the Illumina iScan Instrument using the Human OmniExpress genotyping array (Illumina Inc, San Diego, CA). For each whole genome shotgun library, 500 ng of genomic DNA was fragmented in 5X DNATerminator End Repair Buffer (Lucigen, Middleton, WI), using the Covaris S2 and micro-TUBEs (Covaris, Woburn, MA), with the following settings: volume 50 µl, temperature 4°C, duty cycle 5, intensity 4, cycle burst 200, time 90 s. The fragmented ends were converted to blunt ends by adding DNATerminator End Repair Enzyme. The blunt-ended DNA was then purified using the MinElute PCR purification kit (Qiagen, Germany). DNA was eluted with 32 µl 10 mM Tris-HCl (pH 8.5). A 3′ A overhang was added to the blunt-ended fragments by treating with 15 units of Klenow Fragment 3′->5′exo- and 200 nM dNTP mix (New England BioLabs, Ipswich, MA) for 30 min at 37°C. We purified each adenylation reaction using MinElute PCR purification kit and DNA was eluted with 20 µl 10 mM Tris-HCl (pH 8.5). Each sample was then ligated with 2.5 µl of a 4 µM stock of Illumina Paired End Adapters. The ligation reactions were accomplished using 5000 units of T4 DNA Quick Ligase (New England BioLabs, Ipswich, MA) and incubated for 15 min at 25°C. Each ligated sample was purified using MinElute PCR purification kit and DNA was eluted in 10 µl 10 mM Tris-HCl (pH 8.0). We PCR-amplified our ligated libraries with minor modifications of the Illumina standard protocol.

To prevent excessive over-amplification, we cycle optimized our libraries. Each 50-µl PCR reaction included 1 µl at 10 ng/µl of ligated DNA, 1X Phusion PCR Master Mix (New England BioLabs, Ipswich, MA), and 200 nM from each Illumina PE 1.0 forward and 2.0 forward reverse primer pair. The reactions were cycled as follows: 98°C 30 s, 98°C 10 s, 65°C 30 s, 72°C 30 s. After cycles 6, 8, 10, and 12 the program was halted and a 5 µl aliquot was collected. Each cycle amplification product was evaluated on a 2.2% agarose Flash Gel (Lonza, Switzerland) and the proper cycle number determined. Ten PCR reactions were amplified at the determined cycle number to enrich for proper adaptor ligated fragments.

Each sample was purified post amplification using the MinElute PCR purification kit. Each library was fractionated on the LabChip XT using the DNA 750 chip (Perkin Elmer, Hopkinton, MA) collecting three unique fractions: 375 bp, 475 bp, and 675 bp with a ±5% covariance. Each library fraction was assessed for concentration and size to determine molarity using the Qubit Fluorometer Quant-iT dsDNA HS assay (Life Technologies, Grand Island NY) and the Agilent BioAnalyzer High Sensitivity DNA Assay (Agilent Technologies, Santa Clara, CA), respectively. The final concentration of each library fraction was verified through qPCR utilizing the KAPA Library Quantification Kit - Illumina/LightCycler® 480 kit (Kapa Biosystems, Woburn, MA) to produce cluster counts appropriate for the Illumina HiSeq2000 platform.

Each genome was loaded on four lanes of the HiSeq2000 version three-flow cell (Illumina, San Diego, CA). 2 x 101 bp read pairs were generated for each sample, yielding approximately 30x sequence coverage for each genome. Each lane of sequence data also underwent CNV analysis. Average haploid coverage was 46.7.

Supplementary Table 1. Sequencing metrics for each individual.

Twin pair	Phenotype	Gender	Haploid coverage	SNP calls	dbSNP concordance
#1	ALS	Male	37.568	3,155,277	98.09
#1	Unaffected	Male	44.049	3,152,397	98.13
#2	ALS	Male	49.125	3,164,595	98.06
#2	Unaffected	Male	50.124	3,148,610	98.06
#3	ALS	Female	49.782	3,203,467	98.01
#3	Unaffected	Female	43.488	3,175,554	98.05
#4	ALS	Female	48.528	3,168,292	98.09
#4	Unaffected	Female	45.059	3,158,264	98.06
#5	ALS	Female	49.431	3,184,229	97.95
#5	Unaffected	Female	49.836	3,179,649	97.99

Sequence alignment and variant calling

Illumina reads passing instrument QC were aligned to the GRCh37-lite reference sequence with BWA (1) v0.5.9; parameters –t 4 –q 5 were passed to the bwa aln command and defaults were used for other commands. Duplicates were marked by Picard v1.46. Single-nucleotide variants were called using VarScan (2) v2.3.6 (with parameters—min-coverage 3 --min-var-freq 0.20 --p-value 0.10 --strand-filter 1 --map-quality 10) and SAMtools v0.1.16, and filtered to remove false-positives as previously described (3). Small insertion/deletion variants (indels) were called by VarScan v2.3.6 with the same parameters and false-positive filtering.

Adequacy of coverage

At The Genome Institute at Washington University we have used 30x sequencing extensively to successfully identify and validate somatic variants (4–6). Furthermore, taking the union of the calls from all three callers reduces coverage biases that might be present in one of the callers. Also, we have submitted these sequences to the database of Genotypes and

Supplementary Table 2. Algorithms, parameters and filters used for identifying discordant single nucleotide variants (SNVs), insertions/deletions (INDELs), copy number variants (CNVs), and other structural variants (SVs).

	Tool	Parameters	Filters
SNVs	samtools r982	mpileup -BuDS	max-mm-qualsum-diff 100, bam-readcount-min-base-quality 15
SNVs	sniper 1.0.2	-F vcf -q 1 -Q 15	bam-readcount-min-base-quality 15, somatic-score-mapping-quality
SNVs	varscan-somatic 2.3.6	–nobaq–version r982	varscan-high-confidence v1, bam-readcount-min-base-quality 15
SNVs	strelka 1.0.11	isSkipDepthFilters = 0	
INDELs	GATK-somatic-indel 5336		bam-readcount-min-base-quality 15
INDELs	Pindel 0.5		pindel-somatic-calls v1, pindel-vaf-filter v1 (variant freq cutoff = 0.08), pindel-read-support-v1
INDELs	varscan-somatic 2.3.6	–nobaq–version r982	varscan-high-confidence indel v1, bam-readcount-min-base-quality 15
INDELs	strelka 1.0.11	isSkipDepthFilters = 0	
CNVs	copycat-somatic	per-read-length, per-library	var-filter-snv v1
SVs	breakdancer 1.4.2	-g -h; a -t -q 10 -d	novo-realign v1, tigra-validation v1
SVs	breakdancer 1.4.2	-g -h; a -t -q 10 -o	tigra-validation v1
SVs	squaredancer 0.1		tigra-validation v1

Phenotypes (dbGaP) so that others can cross-check their findings with our data.

Difficulty of detecting mosaicism

A mosaic mutation present in both the affected twin and the matched control sample (the unaffected twin) will be difficult to detect. However, the principal challenge in somatic mutation calling is to isolate the relatively small number of somatic events from millions of constitutional variants. The vast majority of variants seen in both samples, but with higher frequencies in the affected twin, are likely to be germline variants. SomaticSniper was specifically developed to be robust to variant-supporting reads present in the normal, relative to other somatic callers. In simulations using mixed samples, the power to detect somatic variants is $>90\%$ for tumors with variant allele frequencies of $>35\%$ and normal variant allele frequencies of $<5\%$.

Criteria used to assess credibility of variants

The bam files for each twin were loaded into IGV and visually inspected by trained personnel who handle the manual review for a variety of projects. Variants were scored as follows: S: somatic; A: ambiguous; O: other (not a good call for some other reason); G: germline; LQ: low quality; D: directional (sequencing artifact); V: variant (no coverage in normal, cannot tell whether germline or somatic); T: not a good transcript (missing start or stop codon, pseudogene, etc.). Variants scored as 'S' were considered somatic (discordant), and variants scored as 'G' were taken forward for concordant analysis. We also attempted to validate 'V' calls.

Comparison of the three single nucleotide variant callers

Wang et al. (7) compared MuTect and VarScan2 and found that although MuTect identified more low coverage somatic SNVs, it missed more somatic SNVs with alternate alleles in the matched normal sample. Similarly, Roberts et al. (8) and Xu et al. (9) found differences in various variant callers (GATK, MuTect, VarScan2, SomaticSniper, JoinSNVMix2, and Strelka) with respect to number of sites and sensitivities to noise and low allelic fraction. For these reasons, we chose to take the union of the three callers to avoid any biases from one particular caller.

Comparison of structural variant algorithms

Due to the challenges of detecting structural variants with relatively short sequencing reads, and the fun-

Supplementary Table 3. Numbers of concordant homozygous potentially-damaging (missense/nonsense, splice and non-stop) variants per twin pair, either present or not present (i.e. novel) in dbSNP137.

	Missense/nonsense		Essential splice		Non-stop	
	Not in dbSNP	In dbSNP	Not in dbSNP	In dbSNP	Not in dbSNP	In dbSNP
#1	12	4048	9	1131	0	5
#2	12	3890	7	1109	0	8
#3	15	4017	5	1101	0	11
#4	14	3842	4	1124	0	9
#5	8	3823	3	1084	0	11

Supplementary Table 4. Concordant homozygous recessive variants not found in dbSNP137 (i.e. novel variants) that also do not overlap segmental duplications or RepeatMasked sequence.

Twin pair	Gene	Chr	Position	Ref	Alt	VEP annotation	Variant type
#3	AGBL1	15	86,940,753	A	G	Non-synonymous Coding 2488,798 N/S Polyphen: probably damaging (0.998) SIFT: deleterious (0)	Missense
#2	AL603965.1	10	47,769,462	C	G	Non-synonymous Coding	Missense
#2	MT-ATP6	MT	9115	A	G	Non-synonymous Coding 589,197 I/V Polyphen: benign (0.012)	Missense

damentally different nature of the algorithms, we typically saw only minimal overlap between the structural variant calling tools CopyCat, SquareDancer, and BreakDancer. CopyCat was used for CNV detection only, while BreakDancer and SquareDancer were used for CNVs and inversions/translocations. None of the algorithms identified discordant CNVs (i.e. unique to the affected twin). The union of BreakDancer and SquareDancer structural variant calls was taken forward to the filtering step, but after filtering and review no valid discordant structural variants were found.

Supplementary Table 5 (see the searchable Excel spreadsheet). Concordant, rare, heterozygous, potentially-damaging variants in ALS twins.

Supplementary Table 6. All 42 concordant variants in known ALS susceptibility genes, binned by functional category.

Gene	Non-synonymous	Synonymous	Splice	5′UTR	3′UTR	Intronic
ALS2	1	2	2	0	0	178
ANG	0	1	0	1	0	27
ATXN2	1	1	0	1	0	112
BCL11B	0	1	0	0	1	203
BCL6	2	1	0	0	2	56
C9orf72	0	0	0	1	2	65
CDH13	1	4	1	0	0	4,458
CDH22	0	2	0	1	0	146
CHMP2B	0	2	0	0	1	49
CNTN6	0	4	2	0	0	810
CRIM1	0	2	1	0	2	425
CRYM	0	1	0	1	0	41
DAO	0	0	0	0	1	37
DCTN1	0	1	0	0	0	12
DIAPH3	1	2	0	0	1	980
DOC2B	0	2	0	0	0	54
EWSR1	1	0	0	0	0	35
FEZF2	0	0	0	1	0	2
FIG4	2	1	2	0	2	250
FUS	0	2	0	1	0	4
GRB14	1	0	0	1	2	150
LUM	0	0	0	1	1	14
NEFH	2	6	0	0	3	18
NETO1	2	0	0	1	15	407
OMA1	1	2	0	3	0	153
OPTN	1	1	1	0	1	89
PCP4	0	0	0	0	0	227
PFN1	0	0	0	0	0	4
RAMP3	1	1	0	0	6	99
RNASE2	0	0	0	0	0	1
SETX	4	4	1	0	4	237
SIGMAR1	1	0	0	0	1	1
SOD1	0	0	0	0	0	10
SOX5	0	0	0	0	1	794
SPG11	2	1	0	0	0	33
SQSTM1	0	2	0	0	3	44
SYT9	2	2	1	0	4	651
TAF15	0	1	0	0	0	37
TARDBP	0	0	0	0	0	16
UBQLN2	0	0	0	0	0	0
VAPB	0	0	0	0	9	107
VCP	0	0	2	0	1	14

Supplementary Table 7. Features of the two concordant rare heterozygous missense variants, both listed in ALSoD, identified in two separate twin pairs.

Twin pair	Gene	Chr	Position	Ref	Alt	dbSNP137 ID	VEP annotation	Variant type	Allele frequencies
#3	SYT9	11	7,437,285	C	G	rs117876446	Non-synonymous Coding 1294,353 L/V Polyphen: probably damaging (0.945) SIFT: deleterious (0)	Missense	C: 99.31% (8004/8060) G: 0.68% (55/8060)
#4	EWSR1	22	29,693,915	G	A	rs41311143	Non-synonymous Coding 1429,470 G/S Polyphen: unknown (0) SIFT: deleterious (0.03)	Missense	A: 0.80% (64/8014) G: 99.20% (7950/8014)

Advantages of whole genome versus exome sequencing

We performed whole genome sequencing to be able to look for discordant structural variants (in addition to single nucleotide and indel variants). Whole genome sequencing provides more complete and uniform coverage of the exome than contemporary exome products, which achieve coverage of only 90–95% of coding bases. We tiered our initial analysis to look at coding and regulatory mutations. We did not further explore the non-coding, non-regulatory space since there were thousands of candidates, and it was not feasible to undertake manual review of all of these. We have, however, deposited the sequence data into dbGaP so that other groups can explore these intronic and intergenic variants.

No validation of concordant variants undertaken

Based on Bayes theorem, the positive predictive value (PPV) of the pipeline depends on the number of mutations in the sample. If we expect a large number of mutations, the PPV will be higher than if we expect a small number of mutations. Validation rates from The Genome Institute tumor projects vary from 20% to $>$90% depending on the tumor type (high rates for melanoma and low rates for acute myeloid leukaemia) (4–6). A recent internal analysis of this pipeline using a mixture of samples NA12878 (child) and NA12892 (mother) from the Real Time Genomics call set of NA12878, and her parents from the CEPH/Utah pedigree 1463, suggests that the PPV is expected to max out at around 68–74%, depending on the variant allele fraction. In the case of monozygotic twins, the expectation is that there will be a small number of mutations. With a human de novo single nucleotide variation rate of 1.2×10^{-8} (10) and an estimated rate of early postzygotic de novo SNV rate of $0.04\text{-}0.34 \times 10^{-8}$ per twin pair (11), we would not expect as many mutations as in a typical tumor/normal sample (depending on the type of tumor); hence, the PPV will be low. For the germline variants, we also performed an internal evaluation of the pipeline using the sample NA12878 gold standard set and found a sensitivity of 96.8% and a PPV of 93.1% for all SNVs. For variants with global minor allele frequency of $>$0.1% we have 100% sensitivity and 99.8% PPV. In summary, these figures show that it would be extremely unlikely to obtain false-positives in exactly the same gene location in both twin siblings. We therefore chose not to validate the more than 2000 concordant variants since most of them were in dbSNP at appreciable global minor allele frequency.

Accuracy of manual review

That none of the discordant variants did not pass manual review is not a reflection on the accuracy of this review (no automated methods of review were used). The Genome Center manual reviewers have visualized tens of thousands of mutations over the past several years. They are trained to remove mutations that are obvious artifacts of short-read sequencing and alignment, e.g. local alignment errors, repetitive sequences, and systematic mis-mapping of paralogous sequences. In other words, they remove only clearly false-positives, while mutations that look truly somatic, or which are ambiguous, are retained for subsequent validation. This practice of visually reviewing mutations has been widely adopted for both somatic mutations in cancer (12–16) and for potential disease-causing mutations in rare diseases (17,18).

In our past efforts (i.e. tumor genome sequencing) when considerable numbers of somatic mutations are detected, manual review has proven a valuable screening tool to remove certain kinds of sequencing artifacts. From our experience with manual review in tumor genome sequencing we know that the validation rate tends to be lower when the expected number of true mutations is small (e.g. in leukemia and pediatric tumors) compared to the validation rate for highly mutated tumors (e.g. in lung adenocarcinoma).

In the present study, that some variants passed review but failed to validate is likely a limitation of the sequencing technology and mapping algorithms. Somatic variant calling is inherently prone to false-positives because it relies upon a reference genome assembly that is imperfect. Alignment errors are likely to account for this large proportion of the false-positives. Other factors include undercalling in one or the other twin (e.g. the 'V' calls from manual review) that leads to a false-positive variant call.

Supplementary Table 8. Consensus callset (from the intersection of CNVnator and forestCV calls) of 91 concordant copy number variants (CNVs) that overlapped with genes. Presence in the Database of Genomic Variants (DGV) with 1 = present for loss, gain, and both gain and loss, and overlap with segmental duplications (SegDup = 1) are noted. Genes with asterisks are within novel copy number deletions (see also Supplementary Table 9).

Genes	Chr	Start	End	CNV	DGV Loss	DGV Gain	DGV Both	SegDup	Twin Pairs
ABCA13	7	48653101	48654900	DEL	1	1	0	0	#4
ACOT1	14	73998051	74025000	DEL	1	1	1	1	#2 #4
ADAM3A ADAM5P	8	39232001	39387000	DEL	1	1	1	1	#1 #2 #3 #5
ALPK2*	18	56243001	56264450	DEL	0	0	0	0	#3
ANKRD36	2	97854501	97856500	DEL	1	1	1	1	#1 #3
ANKRD36	2	97860501	97870750	DUP	1	1	1	1	#1 #3 #5
ANKRD36	2	97874001	97901450	DUP	1	1	1	1	#3
ANKRD36B	2	98140151	98158350	DEL	1	1	1	1	#2 #4
ANKRD36BP2	2	89070751	89084000	DUP	1	1	0	1	#1
ASCL3	11	8958001	8965000	DEL	1	1	1	0	#5
BTNL3 BTNL8	5	180375701	180430700	DEL	1	1	1	1	#4
C15ORF29	15	34502051	34505500	DEL	0	0	1	0	#2
C1ORF186	1	206283051	206288500	DUP	1	1	0	0	#1 #4
CCDC144B	17	18507151	18521500	DUP	0	1	1	1	#4
CES1	16	55842501	55865350	DUP	0	1	1	1	#4
CES1P1	16	55796501	55821650	DEL	1	1	1	1	#1
CHEK2P2	15	20392001	20531850	DUP	1	1	1	1	#1
DHRS4L2	14	24450051	24468000	DEL	1	1	1	1	#5
DUSP22	6	293751	379400	DUP	1	1	1	0	#2 #3
DUT	15	48625251	48639150	DUP	0	0	1	0	#5
EEF1DP3	13	32533001	32539000	DEL	1	0	1	0	#4
FAM115A	7	143542401	143549500	DUP	1	1	1	1	#3
FAM149A	4	187093501	187098000	DEL	1	1	0	0	#2 #3
FBN2*	5	127782001	127784000	DEL	0	0	0	0	#2
GBP3	1	89476001	89478500	DEL	1	1	1	0	#3
GHR	5	42628501	42631000	DEL	1	0	0	0	#5
GPRIN2 LOC643650 LOC728643 PPYR1	10	46993501	47150500	DUP	1	1	1	1	#1 #2 #3 #4
GUCY2GP	10	114112101	114116500	DEL	1	1	1	0	#1 #3 #4
GUSBP1	5	21478651	21496950	DUP	1	1	1	1	#2 #5
GUSBP11	22	24025551	24027500	DEL	1	1	0	0	#4
HCG4B HLA-H	6	29851151	29906000	DEL	1	1	1	1	#2 #4
HERC2P3	15	20589501	20628000	DUP	1	1	1	1	#3 #5
HLA-DRB5	6	32454501	32516000	DEL	1	1	1	1	#3 #4 #5
IFNA10 IFNA16 IFNA17 IFNA4 IFNA7	9	21181001	21228000	DEL	1	1	1	1	#4
IL4R	16	27336501	27351000	DEL	1	0	1	0	#2
IMMP2L	7	111097501	111139000	DEL	1	1	1	1	#1
KCNJ12	17	21322901	21354000	DUP	1	1	1	1	#1 #2 #3 #4 #5
LCE1D LCE1E	1	152760501	152770500	DEL	1	1	1	1	#2 #4 #5
LCE3B LCE3C	1	152555501	152588150	DEL	1	1	1	0	#1 #2 #3 #4 #5
LOC100506776	7	38388501	38397500	DEL	1	1	1	1	#4 #5
LOC100506990	8	12395201	12456000	DUP	1	1	1	1	#1 #2 #3 #4 #5
LOC388692	1	149258051	149328000	DUP	1	1	1	1	#1 #2 #3 #5
LOC390705 LOC729264	16	32241101	32662200	DUP	1	1	1	1	#1 #2 #3 #5
LOC440434 TBC1D3 TBC1D3F	17	36330751	36405950	DUP	1	1	1	1	#2 #3 #4
LOC642236	9	68377501	68440000	DUP	1	1	1	1	#1 #5
LOC646214	15	21900001	21942000	DUP	1	1	1	1	#3 #5
LOC727924	15	22294501	22336850	DUP	1	1	1	1	#2
MACROD2-AS1	20	14720301	14974500	DEL	1	1	1	0	#5
MIR570 MUC20 SDHAP2	3	195406001	195450400	DUP	1	1	1	1	#1
MRGPRX1	11	18943651	18964400	DUP	1	1	1	0	#2
NARS2*	11	78188001	78197450	DEL	0	0	0	0	#5
NBEAP1	15	20846251	20878500	DUP	1	1	1	1	#3 #5
NBPF10 NOTCH2NL	1	145138001	145318500	DUP	1	1	1	1	#1 #2 #3
NBPF9 PDE4DIP	1	144820351	144896350	DUP	1	1	1	1	#1 #2 #3
NOTCH2	1	120584601	120637450	DUP	1	1	1	1	#3
OR2T11	1	248764951	248798350	DEL	1	1	1	1	#5
OR4N4	15	22373001	22384000	DEL	1	1	1	1	#3
OR51A2 OR51A4	11	4968001	4976700	DEL	1	1	1	1	#2 #4

(*Continued*)

Table 8. (Continued)

Genes	Chr	Start	End	CNV	DGV Loss	DGV Gain	DGV Both	SegDup	Twin Pairs
OR52N1 OR52N5	11	5784501	5809500	DEL	1	1	1	0	#3
OR5P2	11	7811501	7833500	DEL	1	1	1	0	#5
PCDHA10 PCDHA8 PCDHA9	5	140222501	140239000	DEL	1	1	1	1	#5
PCDHB7 PCDHB8	5	140554401	140558950	DUP	1	1	1	1	#1
PDE4DIP	1	144979501	145082550	DUP	1	1	1	1	#2
PDPR	16	70155501	70201700	DUP	1	1	1	1	#4 #5
PMS2CL	7	6785001	6787500	DEL	1	1	1	1	#2
PRAMEF1 PRAMEF11	1	12856501	12890750	DEL	1	1	1	1	#1
PRB1 PRB2	12	11506601	11545300	DUP	1	1	1	1	#1
PRIM2	6	57205551	57284900	DUP	1	1	1	0	#1 #2 #3 #4 #5
PRIM2	6	57322651	57423000	DUP	1	1	1	0	#1 #2 #4 #5
PRIM2	6	57429101	57533050	DUP	1	1	1	0	#1 #2 #4 #5
PRSS2 TRY6	7	142476001	142486500	DEL	1	0	1	1	#1 #3 #4
PSG1 PSG11 PSG6 PSG7	19	43366651	43546800	DEL	1	1	1	1	#3
RGPD4	2	108443501	108446000	DEL	1	1	1	1	#2 #3
RHCE	1	25732201	25735750	DEL	1	1	1	1	#2
RHD	1	25592651	25610000	DEL	1	1	1	1	#2
RHD	1	25614501	25650800	DEL	1	1	1	1	#2 #3
SIGLEC14 SIGLEC5	19	52133501	52149250	DEL	1	1	1	1	#2
SIRPB1	20	1556251	1588400	DEL	1	1	1	1	#3
SLC25A24	1	108733501	108737500	DEL	1	1	1	0	#2 #5
SLC5A11	16	24916501	24918000	DEL	1	0	0	0	#2
SPINK14	5	147553001	147554300	DEL	1	1	1	0	#1 #2
TAG	5	12675501	12741500	DEL	1	1	1	0	#5
TAS2R43	12	11221501	11250750	DEL	1	1	1	1	#4
TRIM48	11	55032001	55038500	DEL	1	1	1	1	#1 #4 #5
UGT2B17	4	69423151	69490750	DEL	1	1	1	1	#5
UGT2B28	4	70140051	70232000	DEL	1	1	1	1	#5
UPK3B	7	76145001	76163500	DEL	1	1	1	1	#4
ZAN	7	100327501	100340800	DEL	1	0	1	0	#2
ZNF557	19	7086001	7099000	DEL	0	1	1	0	#5
ZNF595 ZNF718	4	14151	68800	DUP	1	1	1	1	#1 #2 #3 #4 #5
ZNF880	19	52889001	52891500	DEL	1	1	0	0	#2

Supplementary Table 9. Seven novel concordant copy number deletions (not present in the Database of Genomic Variants and not overlapping segmental duplications) were identified in the twins. Three of these deletions overlap genes, though none of these is a known ALS susceptibility gene.

	CNV	Start	End	Twin pair	Gene
1	Deletion	206,811,601	206,813,000	#5	NA
5	Deletion	90,684,001	90,688,850	#4	NA
5	Deletion	127,782,001	127,784,000	#2	FBN2
11	Deletion	78,188,001	78,197,450	#5	NARS2
12	Deletion	80,065,501	80,070,500	#2	NA
15	Deletion	41,838,501	41,840,500	#2	NA
18	Deletion	56,243,001	56,264,450	#3	ALPK2

CNV: copy number variant; NA: not applicable.

References

1. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26: 589–95.
2. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25:2283–5.
3. Service SK, Teslovich TM, Fuchsberger C, Ramensky V, Yajnik P, Koboldt DC, et al. Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. PLoS Genet. 2014;10:e1004147.
4. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature. 2010;464:999–1005.
5. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole genome sequencing. Nature. 2012;481:506–10.
6. Ding L, Kim M, Kanchi KL, Dees ND, Lu C, Griffith M, et al. Clonal architectures and driver mutations in metastatic melanomas. PLoS One. 2014;9:e111153.
7. Wang Q, Jia P, Li F, Chen H, Ji H, Hucks D, et al. Detecting somatic point mutations in cancer genome sequencing data: a comparison of mutation callers. Genome Med. 2013;5:91.
8. Roberts ND, Kortschak RD, Parker WT, Seiber AW, Branford S, Scott HS, et al. A comparative analysis of algorithms for somatic SNV detection in cancer. Bioinformatics. 2013;29:2223–30.
9. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. BMC Genomics. 2014;15:244.
10. Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han LD, et al. Estimating the human mutation rate using autozygosity in a founder population. Nat Genet. 2012;44:1277–81.
11. Dal GM, Erguner B, Sagiroglu MS, Yuksel B, Onat OE, Alkan C, et al. Early post zygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. J Med Genet. 2014;51:455–9.

12. Voss MH, Hakimi AA, Pham CG, Brannon AR, Chen YB, Cunha LF, et al. Tumour genetic analyses of patients with metastatic renal cell carcinoma and extended benefit from mTOR inhibitor therapy. Clin Cancer Res. 2014;20: 1955–64.
13. Chang VY, Basso G, Sakamoto KM, Nelson SF. Identification of somatic and germline mutations using whole exome sequencing of congenital acute lymphoblastic leukaemia. BMC Cancer. 2013;13:55.
14. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, et al. Integrated analysis of germline and somatic variants in ovarian cancer. Nat Commun. 2014;5:3156.
15. van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, et al. Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. Cancer Discov. 2014;4:1140–53.
16. Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, et al. RNF43 is frequently mutated in colorectal and endometrial cancers. Nat Genet. 2014;46:1264–6.
17. Mackay DS, Bennett TM, Culican SM, Shiels A. Exome sequencing identifies novel and recurrent mutations in GJA8 and CRYGD associated with inherited cataract. Hum Genomics. 2014;8:19.
18. McInerney-Leo AM, Marshall MS, Gardiner B, Coucke PJ, van Laer L, Loeys BL, et al. Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. Bonekey Rep. 2013;2:456.