

## Supplemental Data

### Haploinsufficiency of *KMT2B*, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia

Michael Zech, Sylvia Boesch, Esther M. Maier, Ingo Borggraefe, Katharina Vill, Franco Laccone, Veronika Pilshofer, Andres Ceballos-Baumann, Bader Alhaddad, Riccardo Berutti, Werner Poewe, Tobias B. Haack, Bernhard Haslinger, Tim M. Strom, and Juliane Winkelmann

**Table S1** Trio-whole-exome sequencing statistics for index family F1

Sample #	Reads	Mapped reads	Percent	Mapped sequence (Gb)	Target bases				Average coverage
					> 1x	> 4x	> 8x	> 20x	
F1-II-5 (index subject)	123,825,063	123,550,183	99.78	12.51	99.78	99.66	99.52	98.47	147.22
F1-I-1 (father)	171,161,611	170,402,003	99.56	17.29	99.92	99.85	99.8	99.2	193.94
F1-I-2 (mother)	145,562,251	145,140,930	99.71	14.7	99.83	99.72	99.65	98.94	168.71

**Table S2** Biallelic, disease-segregating variants identified in index subject F1-II-5

Gene	OMIM disease-association	Genomic position (hg 19)	RefSeq transcript	Variation nucleotide	Variation amino acid	Variant type	Known disease-causing mutation (ClinVar)	dbSNP142	CADD prediction	Frequency in-house exomes <sup>a</sup> (allele count/ total allele number)	Frequency ExAC (allele count/ total allele number)	Mode	Segregation (Sanger sequencing validation)
<i>ANKRD1</i>	no	chr10:92,679,979	NM_014391.2	c.154C>G	p.Pro52Ala	missense	no	rs397517248	0.003	3/15800	13/121400	hom	yes
<i>ABCC2</i>	yes (MIM: 237500)	chr10:101,565,180	NM_000392.4	c.1506G>C	p.Met502Ile	missense	no	not found	29.4	not found	1/121366	hom	yes

Parent-child trio whole-exome sequencing data were used to identify rare (minor allele frequency  $\leq 0.001$ ) protein-altering (including missense, nonsense, splice-site, stop-loss, in-frame insertion and deletion, and frameshift) homozygous and compound heterozygous variants in index subject F1-II-5. After co-segregation testing in all available members of family F1, homozygous missense variants in *ANKRD1* and *ABCC2* were retained. Single heterozygous variants in *ANKRD1* have been linked to cardiomyopathies, without any definitive disease-association reported in the OMIM database. Given the predominant relevance of *ANKRD1* in cardiac and skeletal muscle and its association with cardiac disease, we considered this gene unlikely to be causative for the observed dystonia phenotype. Biallelic mutations of *ABCC2* have been causally related to Dubin-Johnson syndrome (MIM: 237500), a benign congenital liver disease characterized by conjugated hyperbilirubinemia. Given the well-established and exclusive association of *ABCC2* mutations with Dubin-Johnson syndrome, we considered this gene unlikely to be causative for the observed dystonia phenotype. Hom = homozygous. <sup>a</sup>consisting of 7900 non-dystonia exomes.

**Table S3** De novo monoallelic variants identified in index subject F1-II-5

Gene	OMIM disease- association	Genomic position (hg 19)	RefSeq transcript	Variation nucleotide	Variation amino acid	Variant type	Known disease- causing mutation (ClinVar)	dbSNP142	CADD prediction	Frequency in-house exomes <sup>a</sup> (allele count/ total allele number)	Frequency ExAC (allele count/ total allele number)	Read depth (WES)	Variant quality (WES)	Sanger sequencing confirmation
<i>TDRD6</i>	no	chr6:46,655,897	NM_001010870.2	c.32G>T	p.Gly11Val	missense	no	not found	23	not found	not found	39	136	yes
<i>NYNRIN</i>	no	chr14:24,880,350	NM_025081.2	c.2483G>A	p.Gly828Glu	missense	no	not found	24.9	not found	not found	194	225	yes
<i>KMT2B</i>	no	chr19:36,223,856	NM_014727.2	c.6406delC	p.Leu2136Serfs*17	frameshift	no	not found	25.8	not found	not found	90	217	yes

Parent-child trio whole-exome sequencing data were used to identify non-annotated, protein-altering (including missense, nonsense, splice-site, stop-loss, in-frame insertion and deletion, and frameshift) de novo variants in index subject F1-II-5. Sanger sequencing verified the variants and their de novo status. None of the three genes affected by the de novo mutational events were listed as disease-associated in the OMIM database. A single-nucleotide deletion in *KMT2B* was the only variant predicted to have a severe impact on protein structure. WES = whole-exome sequencing. <sup>a</sup>consisting of 7900 non-dystonia exomes.

**Table S4** *KMT2B* loss-of-function (LoF) variants annotated in the ExAC dataset

Genomic position (hg 19)	Reference	Alternate	cDNA change	Protein change	Variant type	Allele count/ total allele number	Predicted protein truncation	QualBy Depth (QD)	Flags
chr19:36,213,260	G	C	c.2458-1G>C	N/A	canonical splice	1/78836	<b>no</b>	12.4	none
chr19:36,214,909	G	T	c.3334+1G>T	N/A	canonical splice	1/106404	<b>no</b>	12.37	none
chr19:36,223,613	G	GC	c.6164dupC	p.Arg2057Profs*3	frameshift	1/92432	yes	4.44	<b>homopolymer run</b>
chr:19:36,223,857	T	TC	c.6408dupC	p.Ala2139Glyfs*6	frameshift	2/70086	yes	<b>1.32</b>	<b>homopolymer run</b>

The ExAC server was queried for high-confidence LoF variant calls (including nonsense, frameshift, and essential splice-site variants) affecting the canonical transcript of *KMT2B* (RefSeq: NM\_014727.2, Ensemble: ENST00000222270). The c.2458-1G>C and c.3334+1G>T splice-site variants are expected not to cause a protein truncation but to maintain the translational reading frame (in-frame skipping of exons 4 and 8, respectively). Annotation for the two listed frameshift variants appears to be dubious and they possibly represent sequencing artefacts as they each map to a homopolymer run of six C nucleotides and show very low QualByDepth (QD) scores.<sup>1</sup> The ExAC-derived probability of being LoF intolerant (pLI) metric for *KMT2B* is 1.0, indicating that it belongs to the set of genes that is highly intolerant to LoF variation.<sup>2</sup>

<sup>1</sup>Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43, 11 10 11-33.

<sup>2</sup>Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285-291.

**Table S5** Phenotypic characteristics of 30 individuals with early-onset generalized dystonia in the replication cohort

Replication cohort (RC) sample #	Sex	Age at sampling (y)	Age at onset (y)	Site of onset	Leg involvement	Generalized dystonia classification	Associated features		Developmental delay	Family history	Trio-WES
							Additional movement abnormalities	Other neurological or systemic manifestations			
RC-1 <sup>a</sup>	F	51	7	hand	+	isolated	none	none		negative	
RC-2 <sup>a</sup>	F	36	8	hand	+	isolated	none	none		negative	
RC-3 <sup>a</sup>	M	29	7	neck	-	isolated	none	none		negative	
RC-4 <sup>a</sup>	F	40	7	neck	+	isolated	none	none		negative	
RC-5 <sup>a</sup>	F	47	5	neck	+	isolated	none	none		negative	
RC-6 <sup>a</sup>	F	63	3	neck	+	isolated	none	none		negative	
RC-7 <sup>a</sup>	M	59	10	hand	+	isolated	none	none		positive	
RC-8 <sup>a</sup>	F	28	2	foot	+	combined	myoclonus	pyramidal signs, hyperreflexia		negative	+
RC-9 <sup>a</sup>	F	43	20	neck	+	isolated	none	none		negative	
RC-10	F	42	29	foot	+	combined	parkinsonism	none		positive	
RC-11	F	57	2	neck	+	isolated	none	none		negative	+
RC-12	M	41	1	hand	+	combined	myoclonus	microcephaly, intellectual disability	+	negative	
RC-13	M	19	0.5	upper limbs	+	isolated	none	none		negative	+
RC-14	M	47	1	neck	+	isolated	none	none		negative	
RC-15	M	20	6	foot	+	combined	myoclonus	microcephaly, cognitive impairment, short stature	+	negative	
RC-16 (F2, F2-II-1)	F	11	3	foot	+	isolated	none	microcephaly, short stature, short weight, strabismus	+	negative	+
RC-17	M	64	28	neck	+	isolated	none	none		negative	
RC-18	M	26	7	hand	+	isolated	none	none		negative	
RC-19	F	33	29	hand	+	isolated	none	none		negative	+
RC-20	F	48	29	foot	+	isolated	none	none		negative	
RC-21	F	50	10	neck	+	isolated	none	none		negative	
RC-22	F	27	6	neck	-	combined	myoclonus	none		negative	
RC-23	F	19	4	neck	+	combined	myoclonus	none		negative	+
RC-24	F	34	4	hand	+	isolated	none	none		positive	
RC-25	F	18	3	hand	+	isolated	none	seizures	+	negative	

RC-26 (F3, F3-II-3)	M	15	11	foot	+	isolated	none	mild intellectual disability, microcephaly, syndactyly, astigmatism	+	negative
RC-27	F	60	17	neck	+	isolated	none	seizures		negative
RC-28	F	29	20	neck	-	isolated	none	none		negative
RC-29 (F4, F4-III-2)	F	6	3	foot	+	isolated	none	mild intellectual disability, microcephaly, short stature, short weight, astigmatism	+	positive
RC-30	M	6	4	hand	+	combined	myoclonus	none	+	negative

<sup>a</sup>previously reported in Zech et al., Mov Disord 2016.<sup>1</sup> The following are the corresponding identifiers from Zech et al., Mov Disord 2016<sup>1</sup> for the replication cohort samples: RC-1=patient #2, RC-2=patient #5, RC-3=patient #6, RC-4=patient #8, RC-5=patient #9, RC-6=patient #10, RC-7=patient #11, RC-8=patient #12, RC-9=patient #15.

<sup>1</sup>Zech, M., Boesch, S., Jochim, A., Weber, S., Meindl, T., Schormair, B., Wieland, T., Lunetta, C., Sansone, V., Messner, M., et al. (2016). Clinical exome sequencing in early-onset generalized dystonia and large-scale resequencing follow-up. Mov Disord. <http://dx.doi.org/10.1002/mds.26808>.

**Table S6** Summary statistics for 30 exomes in the replication cohort

Sample #	Reads	Mapped reads	Percent	Mapped sequence (Gb)	Target bases				Average coverage
					> 1x	> 4x	> 8x	> 20x	
RC-1	102,160,790	102,058,482	99.9	10.32	99.77	99.57	99.25	97.22	128.96
RC-2	112,816,224	112,682,303	99.88	11.39	99.77	99.56	99.27	97.55	144.42
RC-3	136,658,667	136,481,664	99.87	13.8	99.89	99.76	99.58	98.37	165.43
RC-4	132,113,474	131,969,813	99.89	13.34	99.78	99.64	99.6	98.25	161.24
RC-5	97,158,400	97,051,400	99.89	9.81	99.76	99.61	99.4	97.84	123.86
RC-6	96,281,590	96,146,663	99.86	9.72	99.76	99.6	99.36	97.69	122.44
RC-7	134,597,597	134,440,694	99.88	13.59	99.89	99.78	99.67	98.77	164.41
RC-8	116,714,305	116,519,263	99.83	11.79	99.77	99.65	99.51	98.41	140.28
RC-9	111,545,754	111,313,621	99.79	11.27	99.89	99.76	99.58	98.25	136
RC-10	134,803,142	134,230,442	99.58	13.62	99.83	99.7	99.61	98.8	155.89
RC-11	151,491,437	151,059,176	99.71	15.3	99.82	99.71	99.65	98.96	172.98
RC-12	137,054,877	136,157,539	99.35	10.14	99.89	99.79	99.67	98.77	156.84
RC-13	149,145,841	148,597,025	99.63	15.06	99.9	99.81	99.68	99.07	182.13
RC-14	153,037,501	151,847,154	99.22	15.46	99.95	99.82	99.63	98.77	188.67
RC-15	141,958,074	140,692,700	99.11	14.34	99.95	99.81	99.6	98.65	175.66
RC-16 (F2, F2-II-1)	106,695,812	106,088,678	99.43	10.78	99.82	99.63	99.31	97.7	133.78
RC-17	153,647,953	152,409,617	99.19	15.52	99.96	99.84	99.66	98.89	187.43
RC-18	143,675,890	142,482,773	99.17	14.51	99.96	99.84	99.68	98.9	175.98
RC-19	171,499,116	170,136,140	99.21	17.32	99.88	99.75	99.61	99.03	206.32
RC-20	153,587,798	152,801,538	99.49	15.36	99.82	99.71	99.64	98.96	175.96
RC-21	111,612,148	111,612,148	99.88	11.27	99.77	99.64	99.5	98.32	140.71
RC-22	124,013,553	123,157,935	99.31	12.53	99.9	99.75	99.57	98.74	155.62
RC-23	123,741,890	123,583,574	99.87	12.5	99.77	99.66	99.52	98.49	154.64
RC-24	131,020,076	130,861,509	99.88	13.23	99.78	99.67	99.57	98.7	162.02
RC-25	124,368,112	124,002,790	99.71	12.56	99.8	99.67	99.5	98.67	150.05
RC-26 (F3, F3-II-3)	128,583,965	128,408,566	99.86	12.99	99.89	99.78	99.66	98.71	161.41
RC-27	145,640,525	145,374,692	99.82	14.71	99.8	99.65	99.52	98.67	179.75
RC-28	120,865,528	120,721,846	99.88	12.21	99.77	99.66	99.54	98.53	150.01



RC-29 (F4, F4-III-2)	170,878,962	169,272,874	99.06	17.26	99.89	99.76	99.61	99.04	210.64
RC-30	114,051,928	113,905,597	99.87	11.52	99.88	99.75	99.59	98.33	142.17

---

**Table S7** Coverage summary statistics for the target regions of *TDRD6*, *NYNRIN*, and *KMT2B* across 30 exomes in the replication cohort

Gene <sup>a</sup>	Exons <sup>b</sup>									
<i>TDRD6</i> (NM_001010870.2)	1	2	3	4						
	225.2±34.93 (163.42-292.71)	140.45±25.4 (97.9-186.41)	151.05±29.13 (99.75-217.9)	245.88±31.14 (186.68-315.74)						
<i>NYNRIN</i> (NM_025081.2)	1	2	3	4	5	6	7	8		
	131.23±38.35 (71.04-211.12)	243.93±58.66 (160.5-373.01)	231.35±41.14 (156.21-308.34)	189.34±45.15 (127.22-298.89)	253.31±55.9 (148.74-411.82)	98.66±17.36 (70.16-133.98)	218.38±40.82 (147.26-308.66)	195.17±38.42 (130.44-273.04)		
<i>KMT2B</i> (NM_014727.2)	1	2	3	4	5	6	7	8	9	10
	13.85±4.0 (5.07-21.55)	218.19±48.67 (147.41-316.32)	174.3±31.22 (117.55-229.94)	175.4±39.65 (109.77-250.9)	146.42±35.82 (79.53-222.12)	139.45±42.64 (82.32-235.39)	127.11±31.41 (73.56-197.12)	176.1±33.8 (100.21-237.5)	112.42±25.31 (70.44-172.62)	282.88±45.32 (192.47-377.94)
	11	12	13	14	15	16	17	18	19	20
	164.02±42.07 (105.67-237.22)	398.03±87.28 (230.72-530.49)	286.29±58.54 (157.66-374.77)	66.36±16.78 (30.61-95.16)	117.14±31.66 (71.92-179.2)	273.14±56.65 (177.67-404.05)	204.55±42.6 (114.26-289.67)	176.34±33.78 (110.4-231.1)	66.69±17.53 (39.29-100.44)	207.04±46.86 (114.45-282.26)
	21	22	23	24	25	26	27	28	29	30
	306.41±64.14 (188.4-398.22)	276.32±63.84 (161.36-427.48)	119.86±26.06 (74.46-180.41)	207.71±42.31 (137.43-305.51)	216.97±54.33 (124.95-315.4)	111.77±23.61 (69.99-171.22)	197.33±41.11 (116.66-269.47)	132.45±32.21 (71.96-183.51)	129.72±22.9 (77.53-172.32)	90.44±20.59 (56.21-133.21)
	31	32	33	34	35	36	37			
	230.67±41.09 (148.54-313.32)	296.8±56.39 (186.59-379.58)	133.35±33.16 (78.82-183.07)	157.53±30.42 (102.52-218.27)	204.78±47.22 (109.64-309.13)	283.45±62.7 (162.23-408.91)	252.26±70.24 (165.84-410.88)			

<sup>a</sup>candidate gene prioritized in index subject F1-II-5 (NCBI accession number).

<sup>b</sup>for each exon and the adjacent splice junctions, average coverage ± standard deviation and range of coverage (minimum coverage – maximum coverage) across the 30 replication cohort exomes are given.

**Table S8** De novo monoallelic variants identified in individual F2-II-1

Gene	OMIM disease-association	Genomic position (hg 19)	RefSeq transcript	Variation nucleotide	Variation amino acid	Variant type	Known disease-causing mutation (ClinVar)	dbSNP142	CADD prediction	Frequency in-house exomes <sup>a</sup> (allele count/ total allele number)	Frequency ExAC (allele count/ total allele number)	Read depth (WES)	Variant quality (WES)	Sanger sequencing confirmation
<i>HCN2</i>	no	chr19:605,113	NM_001194.3	c.1109T>A	p.Leu370His	missense	no	not found	26.2	not found	not found	19	182	yes
<b><i>KMT2B</i></b>	<b>no</b>	<b>chr19:36,211,882</b>	<b>NM_014727.2</b>	<b>c.1633C&gt;T</b>	<b>p.Arg545*</b>	<b>nonsense</b>	<b>no</b>	<b>not found</b>	<b>36</b>	<b>not found</b>	<b>not found</b>	<b>361</b>	<b>225</b>	<b>yes</b>

Parent-child trio whole-exome sequencing data were used to identify non-annotated, protein-altering (including missense, nonsense, splice-site, stop-loss, in-frame insertion and deletion, and frameshift) de novo variants in individual F2-II-1. Sanger sequencing verified the variants and their de novo status. None of the two genes affected by de novo mutational events were listed as disease-associated in the OMIM database. A nonsense substitution in *KMT2B* was the only variant predicted to have a severe impact on protein structure. WES = whole-exome sequencing. <sup>a</sup>consisting of 7900 non-dystonia exomes.