CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Apollo

1 # Mutations in the Histone Methyltransferase Gene *KMT2B* Cause
2 # Complex Early Onset Dystonia

3 Esther Meyer[1,49], Keren J Carss[2,3,49], Julia Rankin[4,49], John ME Nichols[5,49], Detelina
4 Grozeva[6,7], Agnel P Joseph[8], Niccolo E Mencacci[9], Apostolos Papandreou[1,10],
5 Joanne Ng[1,10], Serena Barral[1], Adeline Ngoh[1,10], Hilla Ben-Pazi[11], Michel A
6 Willemsen[12], David Arkadir[13], Angela Barnicoat[14], Hagai Bergman[15], Sanjay Bhate[10],
7 Amber Boys[16], Niklas Darin[17], Nicola Foulds[18], Nicholas Gutowski[19], Alison Hills[20],
8 Henry Houlden[9], Jane A Hurst[14], Zvi Israel[21], Margaret Kaminska[22], Patricia
9 Limousin[23], Daniel Lumsden[22], Shane McKee[24], Shibalik Misra[25,26], Shekeeb S
10 Mohammed[25,26], Vasiliki Nakou[22], Joost Nicolai[27], Magnus Nilsson[28], Hardev Pall[29],
11 Kathryn J Peall[30], Gregory B Peters[31], Prab Prabhakar[10], Miriam S Reuter[32], Patrick
12 Rump[33], Reeval Segel[34], Margje Sinnema[35], Martin Smith[36], Peter Turnpenny[4],
13 Susan M White[16,37], Dagmar Wieczorek[38,39], Sarah Wiethoff[9], Brian T Wilson[14],
14 Gidon Winter[11], Christopher Wragg[20], Simon Pope[40], Simon JH Heales[40,41],
15 Deborah Morrogh[42], the UK10K Consortium[43], the Deciphering Developmental
16 Disorders Study[43], NIHR Bioresource Rare Diseases Consortium[43], Alan Pittman[9],
17 Lucinda J Carr[10], Belen Perez-Dueñas[44,45], Jean-Pierre Lin[22], Andre Reis[32], William
18 A Gahl[46], Camilo Toro[46], Kailash P Bhatia[23], Nicholas W Wood[9], Erik-Jan
19 Kamsteeg[47], Wui K Chong[48], Paul Gissen[5], Maya Topf[8], Russell C Dale[25,26],
20 Jonathan R Chubb[5], F Lucy Raymond[3,6,7,50] & Manju A Kurian[1,10,50]
21

22 **Affiliations:**

23 1. Molecular Neurosciences, Developmental Neurosciences, UCL-Institute of
24 Child Health, London, UK.

25 2. Department of Haematology, University of Cambridge, NHS Blood and
26 Transplant Centre, Cambridge, UK.
27 3. NIHR BioResource-Rare Diseases, Cambridge University Hospitals,
28 Cambridge Biomedical Campus, Cambridge, UK.
29 4. Clinical Genetics, Royal Devon and Exeter NHS Foundation Trust, Exeter,
30 UK.
31 5. MRC Laboratory for Molecular Cell Biology and Department of Cell and
32 Developmental Biology, UCL, London, UK.
33 6. Department of Medical Genetics, Cambridge Institute for Medical Research,
34 University of Cambridge, Cambridge, UK.
35 7. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
36 8. Institute of Structural and Molecular Biology, Crystallography/Department of
37 Biological Sciences, Birkbeck College, University of London, London, UK.
38 9. Department of Molecular Neuroscience, UCL-Institute of Neurology, London,
39 UK.
40 10. Department of Neurology, Great Ormond Street Hospital, London, UK.
41 11. Pediatric Neurology and Development, Shaare-Zedek Hospital, Jerusalem,
42 Israel.
43 12. Department of Paediatric Neurology, Donders Centre for Brain, Cognition,
44 and Behavior, Radboud University Medical Center, Nijmegen, Netherlands.
45 13. Department of Neurology, Hadassah Medical Center and the Hebrew
46 University, Jerusalem, Israel.
47 14. Department of Clinical Genetics, Great Ormond Street Hospital, London, UK.
48 15. Department of Neurobiology and Neurosurgery, The Hebrew University,
49 Hadassah Medical Centre, Jerusalem, Israel.

50 16. Victoria Clinical Genetics Services, Murdoch Children's Research Institute,
51 Parkville, Victoria, Australia.
52 17. Department of Pediatrics, Institute of Clinical Sciences, University of
53 Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden.
54 18. Department of Clinical Genetics, Southampton General Hospital,
55 Southampton, UK.
56 19. Department of Neurology, Royal Devon and Exeter NHS Foundation Trust,
57 Exeter, UK.
58 20. Bristol Genetics Laboratory, Bristol, UK.
59 21. Functional and Restorative Neurosurgery, Hadassah University Hospital,
60 Jerusalem, Israel.
61 22. Complex Motor Disorders Service, Evelina Children's Hospital, Guy's & St
62 Thomas' NHS Foundation Trust, London, UK.
63 23. Sobell Department, National Hospital for Neurology and Neurosurgery,
64 London, UK.
65 24. Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast,
66 UK.
67 25. Child and Adolescent Health, University of Sydney, Sydney, Australia.
68 26. Institute for Neuroscience and Muscle Research, The Children's Hospital at
69 Westmead, University of Sydney, Sydney, Australia.
70 27. Department of Neurology, Maastricht University Medical Center, Maastricht,
71 Netherlands.
72 28. Department of Pediatrics, Piteå Hospital & Umeå University Hospital, Umeå,
73 Sweden.

74 29. College of Medicine and Dental Studies, The University of Birmingham,
75 Birmingham, UK.
76 30. Neuroscience and Mental Health Research Institute, Institute of
77 Psychological Medicine and Clinical Neurosciences, Cardiff University,
78 Cardiff, UK.
79 31. Department of Cytogenetics, The Children's Hospital at Westmead,
80 Westmead, Australia.
81 32. Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-
82 Nürnberg, Erlangen, Germany.
83 33. Department of Genetics, University of Groningen, University Medical Center
84 Groningen, Groningen, Netherlands.
85 34. Medical Genetics Institute and Pediatrics, Shaare Zedek Medical Center and
86 the Hebrew University School of Medicine, Jerusalem, Israel.
87 35. Department of Clinical Genetics and School for Oncology and
88 Developmental Biology (GROW), Maastricht University Medical Center,
89 Maastricht, Netherlands.
90 36. Department of Paediatric Neurology, John Radcliffe Hospital, Oxford, UK.
91 37. Department of Paediatrics, University of Melbourne, Melbourne, Australia.
92 38. Institute of Human Genetics, University Duisburg-Essen, Essen, Germany.
93 39. Institute of Human Genetics, Heinrich-Heine-University, Medical Faculty,
94 Düsseldorf, Germany.
95 40. Neurometabolic Unit, National Hospital for Neurology and Neurosurgery,
96 Queen Square, London, UK.
97 41. Clinical Chemistry, Great Ormond Street Hospital, NHS Foundation Trust,
98 London, UK.

99 42. North East Thames Regional Genetics Service, Great Ormond Street
100 Hospital, London, UK.
101 43. The members of this consortium/study are listed in the **Supplementary**
102 **Note**.
103 44. Department of Child Neurology, Hospital Sant Joan de Déu, Universitat de
104 Barcelona, Barcelona, Spain.
105 45. Centre for Biomedical Research in Rare Diseases (CIBERER-ISCIII),
106 Hospital Sant Joan de Déu, Barcelona, Spain.
107 46. NIH Undiagnosed Diseases Program, Common Fund, Office of the Director,
108 National Institutes of Health, Bethesda, Maryland, USA.
109 47. Department of Human Genetics, Radboud University Medical Center,
110 Nijmegen, Netherlands.
111 48. Department of Radiology, Great Ormond Street Hospital, London, UK.
112 49. These authors contributed equally to this work.
113 50. These authors contributed equally to this work.
114
115 Correspondence should be addressed to M.A.K. (manju.kurian@ucl.ac.uk)
116

117 **ABSTRACT**

118 Histone lysine methylation, mediated by mixed-lineage leukemia (MLL) proteins, is
119 now known to be critical in the regulation of gene expression, genomic stability, cell
120 cycle and nuclear architecture. Despite being postulated as essential for normal
121 development, little is known about the specific functions of the different MLL lysine
122 methyltransferases. Here we report heterozygous variants in the gene *KMT2B* (also
123 known as *MLL4*) in 27 unrelated individuals with a complex progressive childhood-
124 onset dystonia, often associated with a typical facial appearance and characteristic
125 brain magnetic resonance imaging findings. Over time, the majority of affected
126 individuals developed prominent cervical, cranial and laryngeal dystonia. Marked
127 clinical benefit, including the restoration of independent ambulation in some cases,
128 was observed following deep brain stimulation (DBS). These findings highlight a
129 clinically recognizable and potentially treatable form of genetic dystonia,
130 demonstrating the crucial role of KMT2B in the physiological control of voluntary
131 movement.

132 # INTRODUCTION

133 The control of voluntary movement is governed by interactive neural networks
134 within the brain involving the basal ganglia, sensorimotor cortex, cerebellum and
135 thalamus[1]. Disruption of these pathways can lead to a variety of movement
136 disorders. Dystonia is characterized by sustained or intermittent muscle
137 contractions causing abnormal, often repetitive, movements and postures affecting
138 the limbs, trunk, neck and face. Dystonic movements are typically patterned,
139 twisting, and may be tremulous, often initiated or worsened by voluntary action and
140 associated with overflow muscle activation[2].

141 Dystonia is described in a broad spectrum of genetic and acquired disorders, either
142 in isolation or combined with other neurological and systemic features[1-5]. Despite
143 genetic advances, the underlying cause remains elusive for a significant proportion
144 of individuals with childhood-onset dystonia, hindering future prognostication and
145 treatment strategies[6]. We report 27 individuals with an early-onset, complex,
146 combined progressive dystonia associated with mono-allelic variants in *KMT2B*
147 (*MLL4*, NM_014727.2). *KMT2B* encodes a lysine histone methyltransferase,
148 involved in H3K4 methylation, an important epigenetic modification associated with
149 active gene transcription.

150 # RESULTS

151 ***Chromosomal microdeletions and intragenic KMT2B sequence variants in***
152 ***early-onset dystonia***

153 We identified 34 individuals with undiagnosed childhood-onset dystonia for
154 molecular genetic investigation (**Online Methods, Supplementary Table 1**,
155 **Supplementary Fig. 1**). On routine diagnostic testing, one case (Patient 1) was
156 found to have a microdeletion at 19q13.12 of undetermined significance[7].
157 Diagnostic chromosomal microarray was performed in 23/34 individuals and
158 overlapping microdeletions were detected in a further 5 cases (**Supplementary**
159 **Table 1**, Patients 2-6). Using established external networks (**Online Methods,**
160 **Supplementary Fig. 1**), 4 more cases (Patients 7-10) with microdeletions were
161 identified. In total, 10 patients (Patients 1-10) had overlapping heterozygous
162 interstitial microdeletions at 19q13.11-19q13.12 (**Table 1, Fig. 1a**). Microdeletions
163 detected on diagnostic microarray were verified by established laboratory protocols
164 and confirmed as *de novo* where parental testing was possible (**Supplementary**
165 **Table 2**). The smallest region of overlap extended from 36,191,100-36,229,548bp
166 (GRCh37/hg19), encompassing two HUGO Gene Nomenclature Committee
167 curated genes, *ZBTB32* (zinc finger and BTB domain containing 32) and *KMT2B*
168 (*MLL4*) (**Fig. 1a**).

169 For the remaining 28 cases without a 19q microdeletion, we performed either whole
170 exome (n=6) or genome sequencing (n=9) in 15 (**Online Methods**). Heterozygous
171 sequence variants within *KMT2B* were identified in 6/15 cases (Patients 13, 14, 17,
172 21, 22, 27). Sanger sequencing of *KMT2B* in the other 13 individuals identified one
173 additional mutation-positive case (Patient 16). Through national and international
174 collaborations (**Online Methods, Supplementary Fig. 1**), a further 10 cases

175 (Patients 11, 12, 15, 18, 19, 20, 23, 24, 25, 26a) were subsequently ascertained.
176 Overall, a total of 17 patients with intragenic heterozygous *KMT2B* variants were
177 identified (**Table 1, Fig.1b**). These frameshift insertions (n=1), frameshift deletions
178 (n=6), splice site (n=1), stop-gain (n=2) and missense (n=7) variants were
179 confirmed by Sanger sequencing (**Supplementary Table 2, 3**). Whole exome and
180 genome analysis did not identify pathogenic variants in (i) *ZBTB32*, (ii) known
181 dystonia genes and (iii) genes causing other neurodevelopmental disorders. Where
182 possible, mutations in *TOR1A* (NM_000113.2)*, THAP1* (NM_018105.2) *and GNAL*
183 (NM_182978.3) were excluded by diagnostic single gene testing, next generation
184 multiple gene panels and research Sanger sequencing (**Supplementary Table 4**).

185 Parental DNA was available for 23/27 cases, and familial segregation studies
186 verified that interstitial deletions or intragenic variants had arisen *de novo* in 20
187 patients (**Supplementary Table 2, Supplementary Fig. 2**). Three patients had
188 maternally inherited missense variants (Patient 22, 26a and 27). The *KMT2B*
189 variant identified in Patient 26a had occurred *de novo* in his symptomatic mother
190 (Patient 26b) (**Supplementary Table 2**).

191 ***Phenotypic characterization of patients with KMT2B variants***

192 We identified 27 patients (current age 6-40 years, 14 female, 13 male) with *KMT2B*
193 variants, who presented with clinical symptoms in childhood (**Table 1, Table 2,**
194 **Supplementary Table 5, Supplementary Videos 1-7**). Individuals presenting in
195 early childhood (1-9 years, median age 4 years) had either limb or cranio-cervical
196 symptoms. Clinical presentation for those with microdeletions, frameshift, splice-site
197 and stop-gain mutations occurred significantly earlier (mean age 4.1 years) than for
198 those with intragenic missense variants (mean age 6.4 years) (p-value 0.0223)

199 (**Supplementary Fig. 3**). Most patients (21/27) had lower limb symptoms at
200 disease onset, leading to foot posturing, toe-walking and gait disturbance (**Fig. 2a**).
201 4/27 patients presented initially with upper limb symptoms associated with
202 abnormal postures (**Fig. 2b,c**) and dystonic tremor, leading to reduced dexterity
203 and handwriting difficulties (**Supplementary Fig. 4a,b**). With increasing age,
204 cervical symptoms (torticollis, retrocollis) (**Fig. 2d,e**) and cranial involvement (facial
205 dystonia, oromandibular involvement with dysarthria/anarthria and difficulties in
206 chewing/swallowing) became prominent features in the majority of patients. In many
207 patients, progressively severe dysphonia was suggestive of laryngeal involvement.
208 None of the patients had airway compromise and videostroboscopy was not
209 undertaken. Over time, most patients (24/27) developed progressive, generalized
210 dystonia, 2-11 years after initial presentation (**Fig. 2f**). The dystonia was persistent
211 in nature, absent in sleep, worsened by voluntary action and associated with
212 overflow muscle activation. Some patients had dystonic tremor. Sudden, brief,
213 involuntary muscle jerks, clinically consistent with myoclonus, were evident in 2
214 cases (Patients 14 and 27). For a few subjects, dystonia was exacerbated when
215 systemically unwell. Stepwise deterioration following intercurrent illness was
216 particularly evident in Patient 14, and status dystonicus, triggered by a urinary tract
217 infection, was reported in Patient 3.

218 Many patients with *KMT2B* variants had additional clinical findings, including
219 microcephaly, seizures, spasticity and eye movement abnormalities (strabismus,
220 saccade initiation failure and oculomotor apraxia) (**Table 2**). Dysmorphic features
221 and characteristic facial appearance (elongated face and bulbous nasal tip) (**Fig.
222 2g, Table 2**) were commonly reported. Developmental delay, intellectual disability,
223 systemic (dermatological, renal, respiratory) features and psychiatric symptoms

224 were also present in some individuals (**Table 2, Supplementary Table 5,**
225 **Supplementary Fig. 4c**). Malignancies were not reported in any patients.
226 Cerebrospinal fluid (CSF) neurotransmitter analysis, undertaken in 13 patients
227 revealed no major derangement of monoamine metabolites (**Supplementary Table**
228 **6**). Magnetic resonance (MR) imaging revealed a characteristic signature in 17/22
229 patients who had imaging sequences suitable for assessment (**Supplementary**
230 **Table 7**). Subtle, symmetrical hypointensity of the globus pallidi (with a hypointense
231 streak of bilateral globus pallidus externa) was evident on MR images known to
232 demonstrate the magnetic resonance phenomenon of susceptibility (T2, T2*-,
233 susceptibility- and echo-planar imaging b0-diffusion-imaging datasets) (**Fig. 3**).
234 Mean age at neuroimaging was significantly lower for patients with MR
235 abnormalities (11.7 years) than for those with normal brain scans (19.0 years) (p-
236 value 0.0167) (**Supplementary Fig. 5a-c**). Single positron emission tomography
237 using ^{123}I (DaTSCAN^{TM}) and ^{18}F-FDG-PET-CT glucose uptake studies, each
238 undertaken in 3 patients, were normal (**Supplementary Table 7, Supplementary**
239 **Fig. 5d**).

240 ***Deep brain stimulation: clinical benefit in KMT2B-dystonia***

241 Overall, medical therapies were not clinically beneficial. None of the patients had a
242 sustained response to levodopa treatment, nor other commonly used anti-dystonic
243 agents (**Table 1**). Ten patients had symptomatic treatment with bilateral globus
244 pallidus interna-deep brain stimulation (GPi-DBS) (**Table 1**). All showed clinical
245 benefit, which was particularly striking in some of the younger patients. Patient 6
246 showed significant reduction of torticollis and retrocollis, with improvements in
247 motor function and gait. Patient 8 showed a sustained clinical response 6 years

248 after DBS insertion, with improvement of dystonia, even more evident after
249 replacement of a faulty right DBS lead. Patient 9 had generalized dystonia and
250 could not walk independently prior to DBS. Two weeks after DBS insertion, he
251 dramatically regained independent ambulation with marked improvement of
252 dystonic symptoms (**Supplementary Video 8**). Patient 17 and 21 were
253 predominantly wheelchair-dependent prior to DBS insertion, but both patients
254 showed restoration of independent walking and improvement of dystonia after DBS
255 (**Supplementary Video 9, 10**). Patient 19 had amelioration of oromandibular
256 symptoms with DBS. Patient 20 had DBS inserted at age 32 years and although
257 most benefits were only transient, sustained improvement of foot posture was
258 reported. Patient 23 had significant reduction of dystonic symptoms after DBS
259 insertion. Patient 22, 9 months after DBS insertion (**Supplementary Video 11**) and
260 Patient 25, 4 months after DBS insertion, have both shown significant gains in hand
261 function and independent walking with improvement of dystonia. Five patients are
262 now over three years post-DBS insertion, and all report a sustained reduction in
263 dystonia, with restoration of function and prevention of progressive disability.

264 ***KMT2B is constrained for missense and predicted protein truncating variants***

265 Patient 13, 14, 17 and 21 had whole genome sequencing as part of the NIHR-
266 funded BioResource-Rare Disease project. Enrichment analysis was undertaken in
267 this cohort to determine whether predicted protein truncating variants (PPTVs) in
268 *KMT2B* were observed more frequently in patients than would be expected by
269 chance. Given the size and sequence context of *KMT2B*, 5.73×10^{-03} *de novo*
270 *KMT2B* PPTVs would be expected to occur by chance in the subset of the NIHR
271 BioResource-Rare Diseases cohort with pediatric onset neurological disease, but 3

272 PPTVs were observed. This represents a significant enrichment (p-value 3.12×10^{-08}). Furthermore in ExAC, *KMT2B* is highly constrained for PPTVs (accessed July 2016)[8] providing supportive evidence of its potential involvement in disease. 712 *KMT2B* missense variants are reported in the ExAC database. Most of these are rare, as expected for a cohort of this size, and the median CADD score[9] for these variants is 22.9. The median CADD score for missense variants identified in our KMT2B-dystonia cohort is significantly higher at 29.1 (p-value 0.0001364; **Supplementary Table 3**). Given the size and sequence context of *KMT2B*, 956 missense variants are predicted to occur by chance, suggesting that *KMT2B* may also be constrained for missense variation (z=4.06)[8].

282 ***KMT2B variants are predicted to destabilize protein structure***

283 *In silico* homology modelling studies were undertaken to generate hypotheses regarding the predicted effects of sequence variants on KMT2B (NP_055542.1) structure-function properties (**Supplementary Notes**). Based on Pfam domain assignments, KMT2B has a CXXC zinc finger domain, multiple PHD domains, an F/Y rich N-terminus (FYRN), FYRC (F/Y rich C-terminus) domain and a C-terminal SET domain (**Fig. 4a**). The modelled variants occurred in residues within the PHD-like, FYRN, SET and FYRC-SET linking domains (**Fig. 4b-d**). Evaluation of a number of variants using MAESTRO[10] and DUET[11] suggests a change in the free energy, with a predicted structure destabilizing effect (**Supplementary Notes**).

292 p.Phe1662Leu and p.Gly1652Asp occur within a PHD-like domain (residues 1574-1688), predicted to facilitate interaction with DNA, protein-protein interaction and recognition of methylated/unmethylated lysines[12-14]. Extensive hydrophobic interactions hold the globular structure of this region, which is important for its

296 function[12]. Phe1662 is fully buried at the core, stabilizing the structure of this PHD-
297 like domain while Gly1652 is partially buried (**Fig. 4b,e,f**). Phe1662 is involved in
298 multiple hydrophobic contacts at the core of the PHD domain, and exchange for
299 leucine is predicted to cause loss of contacts at the core (**Fig. 4g**). Gly1652 is
300 located on a loop (**Fig. 4e**) and substitution to aspartic acid is predicted to alter
301 surface charge, with possible effects on the interaction network in the vicinity,
302 involving a positively charged Arg1635 which is part of the helix α3 implicated in
303 DNA binding[12]. Arg1762 and Leu1781 occur in a FYRN domain. FYRN and FYRC
304 regions, particularly common in MLL histone methyltransferases, interact to form a
305 compact structural unit (**Fig. 4c,h**) important in maintaining the active structure[15,16].
306 Arg1762 forms hydrogen bonds with the backbone carboxyls of Arg2463 and
307 Leu2464 of FYRC domain. Substitution of Arg1762 by cysteine is predicted to
308 abolish these contacts and hence contribute to destabilization of FYRC-FYRN
309 association. Leu1781, at the interface between FYRN and FYRC (**Fig. 4h,i**) is
310 surface exposed and involved in backbone hydrogen bonds stabilizing the beta
311 sheet formed together by the two domains. Substitution to proline (p.Leu1781Pro) is
312 predicted to disrupt the backbone hydrogen bond at this position, because it lacks
313 one hydrogen bond donor and its backbone torsion angles are not compatible with
314 that of a beta sheet. This predicts a destabilizing effect on sheet structure,
315 potentially affecting the normal association of FYRN and FYRC domains. Arg2517
316 resides in the region linking FYRC and SET domains, known to bind WDR5, an
317 effector required for trimethylation of histone H3[17], presenting methylated histone
318 H3 substrates to the MLL complex for further methylation[18]. Arg2517 is thought to
319 be involved in a salt-bridge interaction with Asp172 of WDR5 (NP_438172.1) (**Fig.
320 4j**) and Arg2517Trp is predicted to lead to loss of this interaction. Ile2674, Tyr2688

321 and Ile2694 all occur in the catalytic methyltransferase SET domain common to
322 histone lysine methyltransferases. Ile2674 is buried in the hydrophobic core,
323 adjacent to the catalytic site (**Fig. 4d,k**). Substitution to threonine is predicted to
324 lead to loss of contacts at the core of the domain (due to the shorter side chain) and
325 also introduces a buried polar group (**Fig. 4l**). p.Tyr2688Thr occurs at the core of
326 SET domain involving extensive hydrophobic interactions and a hydrogen bond
327 interaction with Ser2661 (**Fig. 4m**). The frameshift mutation p.Tyr2688Thrfs*50,
328 with insertion of 50 additional residues, is predicted to destabilise the core and
329 affect contacts due to the substitution with a shorter non-aromatic side-chain.
330 Ile2694 is involved in the extensive hydrophobic contacts stabilizing the core of this
331 domain. *In silico* analysis predicts that the frameshift mutation p.Ile2694Serfs*44
332 will disrupt the domain fold and affect methyltransferase activity.

333 ***KMT2B is ubiquitously expressed with reduced expression in KMT2B-***
334 ***dystonia***

335 We confirmed widespread *KMT2B* expression in a variety of control fetal and adult
336 human tissues (**Fig. 5a, Supplementary Fig. 6**). Moreover, *KMT2B* is ubiquitously
337 expressed in the brain with higher expression in the cerebellum than in any other
338 region (**Fig. 5b**). We ascertained fibroblasts from all patients consented for
339 research testing (Patients 2, 13, 14, 16, all with microdeletions or PPTVs in *KMT2B*)
340 and detected a statistically significant decrease in fibroblast *KMT2B* expression on
341 quantitative RT-PCR when compared to control fibroblasts (**Fig. 5c**).

342 ***Histone H3K4 methylation is not globally reduced in KMT2B-dystonia***

343 To determine the effect of *KMT2B* variants on methylation of lysine 4 on histone H3
344 (H3K4 methylation), we assayed tri-methylated H3K4 (H3K4me3) and di-

345 methylated H3K4 (H3K4me2). Immunoblotting of histones extracted from fibroblasts
346 of Patients 14 and 16 showed no significant reduction in H3K4me3 or H3K4me2
347 relative to control samples (**Fig. 5d, Supplementary Fig. 7a**). A *Dictyostelium*
348 *discoideum* model was used to test the effect of SET domain variant p.Ile2647Thr
349 on *in vivo* histone methyltransferase activity. The SET domain of KMT2B shares
350 56**%** sequence identity with the *Dictyostelium* orthologue DdSet1, and Ile2647 is
351 conserved (corresponding residue in *Dictyostelium* is Ile1447, XP_636258.1)
352 (**Supplementary Fig. 8f**). DdSet1 is the only H3K4 methyltransferase in
353 *Dictyostelium* and targeted knockout of *DdSet1* (*set1*$^{-}$) results in loss of all
354 methylation at H3K4[19]. We constitutively expressed wild-type DdSet1 (WT-DdSet1)
355 and mutant-DdSet1 (m-DdSet1), both with N-terminal GFP fusions, in *set1*$^{-}$
356 *Dictyostelium* cells and compared the resulting levels of H3K4 methylation.
357 Expression of either GFP-WT-DdSet1 or GFP-mDdSet1 in *set1*$^{-}$ cells resulted in
358 rescue of H3K4 tri-methylation to wild type levels (**Fig. 5e, Supplementary Fig.**
359 **7b,c**).

360 ***Altered gene and protein expression in KMT2B-dystonia***

361 In order to determine whether KMT2B-dystonia is associated with dysregulation of
362 specific genes and proteins, we investigated (i) gene and protein expression
363 profiles for THAP1 and Torsin-1A in cultured patient fibroblasts from Patients 2, 13,
364 14 and 16 and (ii) tyrosine hydroxylase and dopamine 2 receptor (D2R) protein
365 levels in cerebrospinal fluid from Patients 2 and 16 (**Supplementary Notes,**
366 **Supplementary Fig. 9, Supplementary Fig. 10**). We found significantly reduced
367 transcript levels of *THAP1* and *TOR1A* when compared to control fibroblasts
368 (**Supplementary Fig. 11a**). Fibroblast immunoblotting studies showed a statistically

369 significant reduction in THAP1 protein expression in all 4 patients when compared
370 to control samples (**Supplementary Fig. 11b**). A statistically significant reduction in
371 Torsin-1A was evident in Patient 14, though not in other patients (**Supplementary**
372 **Fig. 11c**). CSF immunoblotting revealed significantly reduced levels of dopamine 2
373 receptor (D2R) and increase in tyrosine hydroxylase (TH) levels (**Supplementary**
374 **Fig. 11d**).
375

376
DISCUSSION

377 We report 27 individuals with heterozygous variants in the lysine methyltransferase
378 gene, *KMT2B,* and define a new genetic movement disorder that, importantly, is
379 amenable to treatment with DBS. Using the current classification system[2], KMT2B-
380 dystonia is defined as an inherited autosomal dominant, complex, combined
381 dystonia usually of infantile or childhood-onset. In most patients, the dystonia is
382 persistent and progressive in nature. Most individuals develop 4-limb dystonia with
383 particularly prominent cervical, laryngeal and oromandibular symptoms. Whilst the
384 majority of patients seem to follow this disease trajectory, we also report one young
385 case (Patient 10, age 7 years) with developmental delay and intermittent toe-
386 walking only. Furthermore, we describe atypical cases with mainly oromandibular
387 features (Patient 18) or paroxysmal cervical dystonia (Patient 26a) and relatively
388 little upper or lower limb involvement.

389 For many patients, KMT2B-dystonia is associated with a number of additional
390 clinical features including other neurological symptoms, intellectual disability,
391 psychiatric co-morbidity, dysmorphia, skin lesions and other systemic signs. Given
392 the association with active gene expression, it is conceivable that *KMT2B* variants
393 could account for these additional disease features. For Patients 1-10, other genes

within the 19q microdeletion may also contribute to aspects of their clinical phenotype[20]. *KMT2B* variants therefore cause a complex dystonia, and affected patients should have close surveillance of development during childhood, regular neurology assessments, routine dermatological review and formal neuropsychiatric testing.

In KMT2B-dystonia, the majority of patients had a characteristic pattern on MR imaging, with subtle, low pallidal signal on T2*-, diffusion- and susceptibility-weighted sequences, particularly affecting the lateral aspect of the globus pallidus externa (**Fig. 3**). Genotype did not appear to influence MR findings. However, those with abnormal imaging had scans undertaken at a significantly younger age than those with normal imaging. MR abnormalities may possibly be an age-dependent phenomenon, perhaps becoming less apparent with increasing age, as evident in serial imaging from Patient 22 (**Supplementary Table 7, Supplementary Fig. 5b,c**). The overall significance of these neuroradiological abnormalities remains unclear. Such findings are reminiscent of, but subtly different to, those reported in Neurodegeneration with Brain Iron Accumulation (NBIA) syndromes[21,22]. Similar non-specific features of T2*-weighted hypointensity are increasingly recognized in other neurological conditions, including Huntington's disease, *TUBB4A*-related disorders, GM1 gangliosidosis, alpha-fucosidosis and mitochondriocytopathies.

KMT2B variants were identified in 13/34 (38%) individuals with a relatively homogenous phenotype of early onset progressive dystonia. For externally screened cohorts, detection rates varied from 1.3-30% according to the phenotypic focus of the cohort (**Supplementary Fig. 1**). For cases where *KMT2B* mutations were not detected, it is likely that another etiology accounts for their symptoms. However, it is possible that *KMT2B* mutations may have been missed as (i)

419 single/multiple exon *KMT2B* deletions and duplications may not be detected on
420 microarray, Sanger sequencing and whole exome/genome sequencing and (ii)
421 promoter mutations and intronic *KMT2B* variants may not have been identified by
422 whole exome and Sanger sequencing.

423 The majority of individuals with *KMT2B* variants (Patients 1-20) had either
424 heterozygous interstitial microdeletions leading to *KMT2B* haploinsufficiency or
425 variants predicted to cause protein truncation, protein elongation, splicing defects or
426 nonsense-mediated mRNA decay. The remaining 7 patients (Patients 21-27) had
427 non-synonymous variants of *KMT2B*. Although a degree of caution must be
428 exercised for missense variants, those identified in our cohort are (i) described in
429 patients with a compatible phenotype, (ii) predicted to affect conserved residues
430 within key protein domains for 5/7 cases (**Supplementary Fig. 8, Supplementary**
431 **Fig. 12**) and (iii) predicted by *in silico* tools to be deleterious with a destabilizing
432 effect on protein structure (**Supplementary Table 3**). Initial disease presentation
433 was significantly earlier in those with missense variants (**Supplementary Fig. 3**)
434 though genotype did not seem to influence subsequent rate of symptom evolution,
435 disease severity or DBS response.

436 For the majority of patients, *KMT2B* variants were confirmed as *de novo* where
437 parental testing could be undertaken. In our cohort, 3 patients had missense
438 changes that were maternally inherited (Patient 22, 26a, 27). The possibility of
439 imprinting at the disease locus was considered, but deemed unlikely, given that (i)
440 *de novo* microdeletions in Patients 2 and 10 occurred on paternally inherited alleles
441 and (ii) there is bi-allelic expression of *KMT2B* single nucleotide polymorphisms in
442 human tissues, including brain (**Supplementary Fig. 13**). Importantly, whole exome
443 sequence analysis undertaken in Patients 22, 26a and 27 did not identify other rare

444 or *de novo* variants to account for disease. Interestingly, Patient 26a inherited
445 p.Arg2517Trp from his symptomatic mother (26b) in whom the change occurred *de*
446 *novo* (**Supplementary Fig. 2**). She was more mildly affected, with onset of
447 symptoms in early adulthood, reporting gait abnormalities, progressive inability to
448 run and periodic paroxysmal upper limb and neck dystonia. Both had similar facial
449 appearances to others in the cohort (**Fig. 2g**). In contrast, the mothers of Patients
450 22 and 27 were clinically examined and neither had evidence of a motor phenotype,
451 intellectual disability, other neurological features, neuropsychiatric symptoms, facial
452 dysmorphia, skin lesions or other systemic signs. The identification of both
453 symptomatic and asymptomatic carriers suggests either 'apparent' incomplete
454 penetrance, due to parental mosaicism, or true incomplete disease penetrance, a
455 phenomenon commonly reported in other autosomal dominant genetic
456 dystonias[23,24]. Other genetic, epigenetic and environmental modifiers may also
457 influence disease penetrance and phenotypic presentation in KMT2B-dystonia.

458 *KMT2B* encodes a ubiquitously expressed lysine methyltransferase specifically
459 involved in H3K4 methylation[25,26], an important epigenetic modification associated
460 with active transcription. H3K4me3 is enriched at promoters, marking transcription
461 start sites of actively transcribed genes, whereas H3K4me1 is associated with
462 active enhancer sequences[27]. H3K4me2 is less specifically localized, but may be
463 enriched at transcription factor binding sites[28]. Members of the SET/MLL protein
464 family, including KMT2B, are responsible for the generation of H3K4me1,
465 H3K4me2, and H3K4me3 which are essential for gene activation in normal
466 development[29]. Using patient-derived fibroblasts and a *Dictyostelium discoideum*
467 model, we demonstrated that *KMT2B* variants are not associated with widespread
468 alterations in overall levels of H3K4 methylation. This is not surprising, given that

469 haploinsufficiency of other MLL family members have not been convincingly shown
470 to affect global H3K4 levels. The fundamental physiological role of MLL proteins is
471 affirmed by the observation that loss-of-function heterozygous mutations in MLL-
472 encoding genes are reported in a number of human developmental disorders[30],
473 namely Wiedemann Steiner (*KMT2A, MLL1*)[31], Kleefstra-like (*KMT2C, MLL3*)[32] and
474 Kabuki (*KMT2D, MLL2*)[33] syndromes, and most recently *SETD1A*-related disease
475 (*KMT2F*)[34]. Although physiological functions of MLL proteins are yet to be fully
476 characterized, the observation that mutations in different *MLL* genes cause
477 phenotypically distinct syndromes (**Supplementary Table 8**) suggests that each
478 MLL protein has a unique role, regulating the expression of a specific set of
479 genes[35,36].

480 Amongst the previously reported *MLL*-gene disorders, dystonia appears fairly
481 specific to KMT2B-related disease and is not commonly described in other MLL
482 syndromes (**Supplementary Table 8**), providing further evidence that different MLL
483 proteins mediate the activation and transcription of a specific set of genes, with
484 temporal and cellular context[37]. In order to determine downstream effects of *KMT2B*
485 mutations, we investigated expression profiles of specific genes and proteins
486 implicated in the pathogenesis of dystonia using patient-derived fibroblasts and
487 CSF (**Supplementary Notes; Supplementary Fig. 9-11**). We detected a
488 statistically significant reduction of *THAP1* and *TOR1A* gene expression and
489 decreased THAP1 protein expression in fibroblasts. CSF immunoblotting studies
490 revealed reduction of D2R protein and increase in TH levels in two patients with
491 KMT2B-dystonia when compared to control CSF samples. The mechanisms
492 causing such alterations in KMT2B-dystonia remain yet to be elucidated. Whilst
493 H3K4 methylation is clearly associated with the process of active transcription,

494 several studies have shown that H3K4 methylation is required, not for absolute
495 transcriptional output, but rather for transcription stability or consistency[38,39], so the
496 effects of *KMT2B* haploinsufficiency could conceivably operate via an intermediary
497 sensitive to stochastic fluctuations. It is highly likely that dysregulation of other
498 genes and proteins are also involved in the disease pathophysiology of KMT2B-
499 dystonia. Further studies will determine whether expression profiles of other genes
500 and proteins are affected in KMT2B-dystonia and contributory to the phenotype.

501 In conclusion, we report *KMT2B* variants in 27 patients with a clinically recognizable
502 form of dystonia. To date, the underlying genetic etiology is only resolved in a
503 minority of childhood-onset cases of dystonia, which precludes confirmatory
504 diagnosis, accurate disease prognostication and selection of appropriate treatment
505 strategies. We have shown that many patients with KMT2B-dystonia have
506 significant, sustained clinical improvement with DBS. Referral for DBS assessment
507 should therefore be considered for this group. Identification of additional cases will
508 allow further characterization of the full phenotypic disease spectrum. Our report
509 highlights mutations in *KMT2B* as a new and important cause of complex early-
510 onset dystonia, emphasizing the crucial role of KMT2B in the control of voluntary
511 movement.

512 **URLs:**

513 Exome Aggregation Consortium (ExAC) database (accessed July 2016)

514 http://exac.broadinstitute.org

515 DECIPHER

516 http://decipher.sanger.ac.uk

517 UK10K Project

518 http://www.uk10k.org

519 Deciphering Developmental Disorders (DDD) study

520 http://www.ddduk.org/

521 1000 Genomes

522 http://browser.1000genomes.org/index.html

523 NHLBI GO Exome Sequencing Project (release 20130513)

524 http://evs.gs.washington.edu/EVS/

525 Ensembl genome browser

526 http://www.ensembl.org/index.html

527 Primer3

528 http://bioinfo.ut.ee/primer3/

529 Chromas Sequencing software

530 http://www.technelysium.com.au/chromas.html

531 Clustal Omega

532 http://www.ebi.ac.uk/Tools/msa/clustalo/

533 SIFT

534 http://sift.jcvi.org/

535 PolyPhen-2

536 http://genetics.bwh.harvard.edu/pph2/

537 Mutation Taster

538 http://www.mutationtaster.org/

539 Combined Annotation Dependent Depletion (CADD)

540 http://cadd.gs.washington.edu/

541 BRAINEAC

542 http://www.braineac.org.

543 Methods:

544 Methods and any associated references are available in the online version of the
545 paper.

546 Accession codes:

547 **Chromosomal microarray data:** Microarray data for Patient 1 (Ref: 326759),
548 Patient 2 (Ref: 326749), Patient 3 (Ref: 326748), Patient 4 (Ref: 326751), Patient 5
549 (Ref: 326750), Patient 6 (Ref: 326752), Patient 7 (Ref: 285035) and Patient 8 (Ref:
550 280902) are deposited in DECIPHER. The data from Patient 1 (Ref: 326759) and
551 Patient 8 (Ref: 280902) is publically available. The remaining patients did not
552 consent for their data to be publicly released.

553 https://decipher.sanger.ac.uk/search?q=326759#consented-patients/results

554 https://decipher.sanger.ac.uk/search?q=280902#consented-patients/results

555 **NIHR BioResource-Rare Diseases (NIHRBR-RD) Study:** Whole genome
556 sequencing data is deposited in the NIHR BioResource Rare Diseases BRIDGE
557 consortium sequencing projects (short name: NIHR-BR-RD). Accession code:
558 EGAS00001001012. Title of dataset: SPEED childhood dystonia KMT2B dataset:
559 EGAD00001002730. Data is deposited for Patient 1 (Ref: EGAR00001314765);

560 Patient 13 (Ref: EGAR00001320121); Patient 14 (Ref: EGAR00001314777);
561 Patient 17 (Ref: EGAR00001314751) and Patient 21 (Ref: EGAR00001314767).
562 https://www.ebi.ac.uk/ega/home
563 **UK10K Project:** UK10K whole exome sequencing data has been deposited under
564 the name UK10K_RARE_FIND. Accession code: EGAS00001000128. Title of
565 dataset: UK10K_RARE_FIND REL-2013-10-31 variant calling: EGAD00001000750
566 Data is deposited for Patients 22 (Ref: UK10K_FIND5536224) and 27 (Ref:
567 UK10K_FIND5536279).
568 https://www.ebi.ac.uk/ega/studies/EGAS00001000128
569 **Deciphering Developmental Disorders (DDD) study:** Exome sequencing data is
570 accessible via the European Genome-phenome Archive (EGA) under accession
571 EGAS00001000775.
572 https://www.ebi.ac.uk/ega/studies/EGAS00001000775
573 **National Institutes of Health, Bethesda; Institute of Human Genetics,**
574 **Erlangen; Radboud University Medical Center, Nijmegen, UCL-Institute of**
575 **Neurology, London:** Whole exome sequencing data has not been deposited since
576 participating patients have not consented for the data to be publicly released.
577

578 ## Note:

579 Any Supplementary Information and Source Data files are available in the online
580 version of the paper.
581

582 ## Acknowledgements:

583 We would like to thank all our patients and their families for taking part in this study
584 and encouraging international collaboration to seek out similar cases. Thank you to

585 Dr Miho Ishida for kindly providing the fetal cDNA, Dr Karin Tuschl for kindly
586 providing the human cDNA panel, Dr Lorenzo Bassioni for kindly selecting
587 DaTSCAN images for the Supplementary manuscript, Dr Matthew Adams for
588 reviewing the imaging of the NHNN patients and Rowdy Meijer for helping with the
589 sequencing analysis at the Department of Human Genetics (Nijmegen). We would
590 like to thanks Professors Gudrun Moore and Philip Stanier for proof-reading the
591 manuscript. Many thanks to Aybeniz Panahian-Jand for excellent administrative
592 support. M.A.K. has a Wellcome Intermediate Clinical Fellowship (WT098524MA).
593 E.M. and M.A.K. received funding from the Rosetrees Trust, Great Ormond Street
594 Hospital Children's Charity and Gracious Heart Foundation. N.E.M. receives
595 support from the Department of Health's National Institute for Health Research
596 Biomedical Research Centers funding streams. A.P. has a joint Action Medical
597 Research/ British Paediatric Neurology Association Research Training Fellowship.
598 J.N. has a Medical Research Council Research Training Fellowship. A.N. has an
599 Action Medical Research Training Fellowship. H.B.P. has a DBS training travel
600 grant from the Daniel Turnberg Trust Fund. H.H. is funded by MRC and Wellcome
601 Trust (Synaptopathies award). D.A. is supported by the Prusiner-Abramsky Award.
602 H.P. has received grant support from The Dystonia Society (UK). K.J.P. has an
603 Academy of Medical Sciences Clinical Starter Grant. B.P.D. received funding from
604 grants 20143130-La Marató de TV3 and PI15/00287-Ministerio Español de
605 Economia y Competitividad. J-P.L. has been supported by a Guy's and St Thomas
606 Charity New Services and Innovation Grant: G060708; The Dystonia Society (UK):
607 Grants 01/2011 and 07/2013 and an Action Medical Research: AMR - GN2097.
608 This research was supported by the National Institute for Health Research
609 Biomedical Research Centre at Great Ormond Street Hospital for Children NHS

610 Foundation Trust, University College London, University of Cambridge and from the
611 NIHR for the BioResource for Rare Diseases (grant number RG65966). This study
612 makes use of data generated by the DECIPHER community. A full list of centres
613 contributing to the generation of the data is available from
614 http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding
615 for the project was provided by the Wellcome Trust for UK10K (WT091310) and
616 DDD Study. The DDD study presents independent research commissioned by the
617 Health Innovation Challenge Fund [grant number HICF-1009-003] - see
618 www.ddduk.org/access.html for full acknowledgement. This work was supported in
619 part by the Intramural Research Program of the National Human Genome Research
620 Institute and the Common Fund, NIH Office of the Director. This work was
621 supported in part by the German Ministry of Research and Education (grant nos.
622 01GS08160 and 01GS08167; German Mental Retardation Network) as part of the
623 National Genome Research Network to A.R. and D.W. and by the Deutsche
624 Forschungsgemeinschaft (AB393/2-2) to A.R. Brain expression data was provided
625 by the UK Human Brain Expression Consortium (UKBEC), which comprises John
626 A. Hardy, Mina Ryten, Michael Weale, Daniah Trabzuni, Adaikalavan Ramasamy,
627 Colin Smith and Robert Walker, affiliated with UCL Institute of Neurology (J.H.,
628 M.R., D.T.), King's College London (M.R., M.W., A.R.) and the University of
629 Edinburgh (C.S., R.W.).

630

631 Author contributions

632 E.M., K.J.C., J.M.E.N., J.R.C., F.L.R. and M.A.K. conceived and designed
633 experiments. J.R., N.E.M., A.P., J.N., H.B-P., M.A.W., D.A., A.Ba., H.B., S.B., N.D.,
634 N.F., N.G., A.H., H.H., J.A.H., Z.I., M.K., P.L., D.L., S.Mc., S.M., S.S.M., V.N., J.Ni.,

635 M.N., H.P., K.J.P., G.B.P., P.P., M.S.R., P.R., R.S., M.Si., M.Sm., P.T., S.M.W.,
636 D.W., B.T.W., G.W., UK10K Consortium, DDD study, NIHRBR-RD study, L.J.C.,
637 B.P-D., J-P.L., A.R., W.A.G., C.T., K.P.B., N.W.W., E-J.K., P.G., R.C.D., F.L.R. and
638 M.A.K. ascertained patients, contributed clinical information, photographs, videos
639 and neuroimaging studies. M.A.K. performed phenotypic characterization of all
640 patients. W.K.C. and M.A.K. reviewed patient neuroimaging. A.P. and M.A.K. edited
641 patient videos. A.Bo., C.W. and D.M. undertook chromosomal microarray analysis.
642 E.M., K.J.C., D.G., N.E.M., S.W., A.Pi., UK10K Consortium, DDD study, NIHRBR-
643 RD study, A.R., W.A.G., C.T., E-J.K. and M.A.K. carried out whole exome/genome
644 sequencing analysis. E.M. and A.N. performed variant validation by direct Sanger
645 Sequencing. K.J.C. performed enrichment analysis (and corresponding statistical
646 analysis). S.P. and S.J.H.H. analyzed CSF neurotransmitters. A.P.J. and M.T.
647 undertook comparative homology modelling. J.M.E.N. and J.R.C. undertook the
648 histone methylation assay (and corresponding statistical analysis) and cloning of
649 Set1 Point Substitution in Dictyostelium. S.B. generated dopaminergic neurons,
650 collected RNA and cDNA samples and undertook quantitative RT-PCR
651 experiments. E.M. maintained fibroblast cultures, collected RNA, cDNA and protein
652 samples, performed fibroblast immunoblotting analysis (and corresponding
653 statistical analysis) and CSF immunoblotting (and corresponding statistical
654 analysis). J.N. carried out CSF immunoblotting analysis. P.G. and F.L.R.
655 contributed critical suggestions for experimental work. E.M. and M.A.K. wrote the
656 manuscript. K.J.C., J.R., J.M.E.N., D.G., A.P.J., N.E.M., A.R., W.A.G., C.T., E-J.K.,
657 W.K.C., M.T., J.R.C. and F.L.R contributed written sections for manuscript. M.A.K.
658 oversaw the overall project. All authors critically reviewed manuscript.
659

660 ## Competing financial interests

661 H.P. has unrestricted support for Educational Activity from Medtronic.

662 ## References

663 1. Charlesworth, G., Bhatia, K.P. & Wood, N.W. The genetics of dystonia: new
664 twists in an old tale. *Brain* **136**, 2017-2037 (2013).
665 2. Albanese, A. *et al.* Phenomenology and classification of dystonia: a
666 consensus update. *Mov. Disord.* **28**, 863-873 (2013).
667 3. Ng, J., Papandreou, A., Heales, S.J. & Kurian, M.A. Monoamine
668 Neurotransmitter Disorders – clinical advances and future perspectives. *Nat.*
669 *Rev. Neurol.* **11**, 567-584 (2015).
670 4. Shanker, V. & Bressman, S.B. Diagnosis and Management of Dystonia.
671 *Continuum* (Minneap Minn). **22**, 1227-1245 (2016).
672 5. Balint, B. & Bhatia, K.P. Isolated and combined dystonia syndromes - an
673 update on new genes and their phenotypes. *Eur. J. Neurol.* **22**, 610-617
674 (2015).
675 6. Lin, J.P., Lumsden, D.E., Gimeno, H. & Kaminska, M. The impact and
676 prognosis for dystonia in childhood including dystonic cerebral palsy: a
677 clinical and demographic tertiary cohort study. *J. Neurol. Neurosurg.*
678 *Psychiatry* **85**, 1239-1244 (2014).
679 7. Dale, R.C., Grattan-Smith, P., Nicholson, M. & Peters, G.B. Microdeletions
680 detected using chromosome microarray in children with suspected genetic
681 movement disorders: a single-centre study. *Dev. Med. Child Neurol.* **54**, 618-
682 623 (2012).
683 8. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans.
684 *Nature* **536**, 285-291 (2016).
685 9. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., & Shendure,
686 J. A general framework for estimating the relative pathogenicity of human

687 genetic variants. *Nat. Genet.* **46**, 310-315 (2014).

688 10. Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S. & Lackner, P. MAESTRO--
689 multi agent stability prediction upon point mutations. *BMC Bioinformatics* **16**,
690 116 (2015).

691 11. Pires, D.E., Ascher, D.B. & Blundell, T.L. DUET: a server for predicting
692 effects of mutations on protein stability using an integrated computational
693 approach. *Nucleic Acids Res.* **42**, W314-319 (2014).

694 12. Liu, Z. *et al.* Structural and functional insights into the human Borjeson-
695 Forssman-Lehmann syndrome-associated protein PHF6. *J. Biol. Chem.* **289**,
696 10069-10083 (2014).

697 13. Musselman, C.A. & Kutateladze, T.G. Handpicking epigenetic marks with
698 PHD fingers. *Nucleic Acids Res.* **39**, 9061-9071 (2011).

699 14. Sanchez, R. & Zhou, M.M. The PHD finger: a versatile epigenome reader.
700 *Trends Biochem. Sci.* **36**, 364-372 (2011).

701 15. Hsieh, J.J., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S.J.
702 Proteolytic cleavage of MLL generates a complex of N- and C-terminal
703 fragments that confers protein stability and subnuclear localization. *Mol. Cell.*
704 *Biol.* **23**, 186-194 (2003).

705 16. Pless, B. *et al.* The heterodimerization domains of MLL-FYRN and FYRC--
706 are potential target structures in t(4;11) leukemia. *Leukemia* **25**, 663-670
707 (2011).

708 17. Wysocka, J. *et al*. WDR5 associates with histone H3 methylated at K4 and is
709 essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859-
710 872 (2005).

711 18. Song, J.J. & Kingston, R.E. WDR5 interacts with mixed lineage leukemia

712 (MLL) protein via the histone H3-binding pocket. *J. Biol. Chem.* **283**, 35258-
713 35264 (2008).
714 19. Chubb, J.R. *et al.* Developmental timing in Dictyostelium is regulated by the
715 Set1 histone methyltransferase. *Dev. Biol.* **292**, 519-532 (2006).
716 20. Malan, V. *et al.* 19q13.11 deletion syndrome: a novel clinically recognisable
717 genetic condition identified by array comparative genomic hybridisation. *J.*
718 *Med. Genet.* **46**, 635-640 (2009).
719 21. Kruer, M.C. *et al.* Neuroimaging features of neurodegeneration with brain
720 iron accumulation. *AJNR Am. J. Neuroradiol.* **33**, 407-414 (2012).
721 22. Meyer, E., Kurian, M.A. & Hayflick, S.J. Neurodegeneration with Brain Iron
722 Accumulation: Genetic Diversity and Pathophysiological Mechanisms. *Annu.*
723 *Rev. Genomics Hum. Genet.* **16**, 257-279 (2015).
724 23. Ozelius, L. *et al.* SourceGeneReviews® [Internet]. Seattle (WA): University of
725 Washington, Seattle; [updated 2014 Jan 02] (1993-2016).
726 24. Klein, C. *et al.* SourceGeneReviews® [Internet]. Seattle (WA): University of
727 Washington, Seattle; [updated 2014 May 1] (1993-2016).
728 25. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705
729 (2007).
730 26. Black, J.C., Van Rechem, C. & Whetstine, J.R. Histone lysine methylation
731 dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**,
732 491-507 (2012).
733 27. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised
734 enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U S A*
735 **107**, 21931-21936 (2010).
736 28. Wang, Y., Li, X. & Hu, H. H3K4me2 reliably defines transcription factor

737 binding regions in different cells. *Genomics* **103**, 222-228 (2014).
738 29. Shao, G.B. *et al.* Dynamic patterns of histone H3 lysine 4 methyltransferases
739 and demethylases during mouse preimplantation development. *In Vitro Cell*
740 *Dev. Biol. Anim.* **50**, 603-613 (2014).
741 30. Shen, E., Shulha, H., Weng, Z. & Akbarian, S. Regulation of histone H3K4
742 methylation in brain development and disease. *Philos. Trans. R. Soc. Lond.*
743 *B. Biol. Sci.* **369** (2014).
744 31. Jones, W.D. *et al.* De novo mutations in MLL cause Wiedemann-Steiner
745 syndrome. *Am. J. Hum. Genet.* **91**, 358-364 (2012).
746 32. Kleefstra, T. *et al.* Disruption of an EHMT1-associated chromatin-
747 modification module causes intellectual disability. *Am. J. Hum. Genet.* **91**,
748 73-82 (2012).
749 33. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of
750 Kabuki syndrome. *Nat. Genet.* **42**, 790-793 (2010).
751 34. Singh, T. *et al.* Rare loss-of-function variants in SETD1A are associated with
752 schizophrenia and developmental disorders. *Nat. Neurosci.* **19**, 571-577
753 (2016).
754 35. Micale, L. *et al.* Molecular analysis, pathogenic mechanisms, and
755 readthrough therapy on a large cohort of Kabuki syndrome patients. *Hum.*
756 *Mutat.* **35**, 841-850 (2014).
757 36. Ang, S.Y. *et al.* KMT2D regulates specific programs in heart development via
758 histone H3 lysine 4 di-methylation. *Development* **143**, 810-821 (2016).
759 37. Jakovcevski, M. *et al.* Neuronal Kmt2a/Mll1 histone methyltransferase is
760 essential for prefrontal synaptic plasticity and working memory. *J. Neurosci.*
761 **35**, 5097-5108 (2015).

762 38. Benayoun B.A., *et al.* H3K4me3 breadth is linked to cell identity and
763 transcriptional consistency. *Cell* **158,** 673-688 (2014).
764 39. Muramoto, T., Müller, I., Thomas, G. Melvin, A. & Chubb, J.R. Methylation of
765 H3K4 Is required for inheritance of active transcriptional states. *Curr. Biol.*
766 **20,** 397-406 (2010).
767 40. Trabzuni, D. *et al*. Quality control parameters on a large dataset of regionally
768 dissected human control brains for whole genome expression studies. *J.*
769 *Neurochem.* **119**, 275-282 (2011).

770 ## Figure legends

771 **Figure 1:**

772 **Molecular Genetics Findings in Patients with *KMT2B* Variants**

773 (**a**) Top panel: Schematic representation of chromosome 19. Middle panel: Ten
774 microdeletions on 19q13.11-19q13.12 (GRCh37/hg19). Lower panel: The smallest region
775 of overlap comprising two genes, *ZBTB32* and *KMT2B*. (**b**) Schematic exon-intron
776 structure of *KMT2B* (NCBI Reference Sequence: NM_014727.2) indicating 7 frameshift
777 insertions and deletions, 2 stop-gain mutations, 1 splice site variant and 7 missense
778 changes.

779

780 **Figure 2:**

781 **Clinical Features of Patients with *KMT2B* Variants**

782 (**a**) Patient 17, age 13 years: gait disturbance with dystonic posturing of the four limbs. (**b**)
783 Patient 27, age 19 years and (**c**) Patient 14, age 18 years: bilateral upper limb dystonic
784 posturing. (**d,e**) Patient 23, age 8 years: retrocollis. (**f**) Patient 12, age 6 years: generalized
785 dystonia, with jaw-opening dystonia and 4-limb posturing. (**g**) Montage of patient faces:
786 Top row (left to right) Patients 1, 2, 3, 4, 8, 9; middle row (left to right) Patients 11, 12, 13,
787 14, 16, 17 and bottom row (left to right) Patients 21, 23, 25, 26a, 26b. Consent to publish
788 patient photographs has been obtained. Facial elongation, broad nasal base and bulbous
789 nasal tip evident in some patients.

790

791 **Figure 3:**

792 **Radiological Features of Patients with KMT2B Variants**

793 Magnetic resonance imaging (MRI) with T2*-weighted (**a,d**) and T2-weighted images (**b,c**),
794 echo-planar technique diffusion-imaging datasets images with b-value of zero (**e-h**) and
795 susceptibility weighted sequences (**i-l**). Abnormal findings indicated by yellow arrows.

796 (**a,e,i**) Representative MRI from control subjects for T2*-weighted sequences (**a**: age
797 10y2m), diffusion-weighted sequences (**e**: age 10y4m) and susceptibility weighted
798 sequences (**i**: age 10y8m) indicating normal appearances of basal ganglia. Patient 1, age
799 9y5m (**b,f,j**), Patient 13, age 11y3m (**c,g,k**), Patient 9, age 15y1m (**d**), Patient 22, age
800 13y1m (**h**) and Patient 25, age 16y (**l**): evidence of bilateral subtle hypointensity of the
801 globus pallidus with hypointense lateral streak of globus pallidus externa.
802
803 **Figure 4:**
804 **Comparative Modelling of KMT2B Protein Structure**
805 (**a**) Schematic domain architecture of KMT2B. (**b-d**) Degree of amino conservation is
806 displayed in the structural models for different domains. Red to blue indicates increasing
807 conservation. (**b**) Model of PHD-like domain shows Gly1652 and Phe1662. (**c**) Model of
808 FYRN domain presents position and conservation of Arg1762 and Leu1781. (**d**) Model of
809 the SET methyltransferase domain indicates position and conservation of Ile2674,
810 Tyr2688 and Ile2694. (**e**) Location of Gly1652 in the PHD-like domain model and the
811 hydrogen bond network in the vicinity (α3 helix is indicated). (**f**) Hydrophobic packing
812 involving Phe1662. (**g**) Change to leucine at 1662 is predicted to cause loss of contacts
813 within the hydrophobic core. Residue side chains are presented as spheres highlighting
814 van der Waals contacts. (**h**) Interactions involving Arg1762 from FYRN with Arg2463 and
815 Leu2464 of FYRC. The hydrogen bond interactions and distances are highlighted. (**i**)
816 Leu1781 shown at the interface of FYRN (orange)/FYRC (magenta) domains. The
817 backbone hydrogen bonds stabilizing the sheet structure are highlighted. (**j**) Interactions
818 involving Arg2517 and WDR5 (beige). The salt bridge interaction between Arg2517 of
819 KMT2B and Asp172 of WDR5 is highlighted. (**k**) Location and contacts involving Ile2674 in
820 the hydrophobic core of the SET domain (SAH is indicated). (**l**) Substitution with threonine
821 at 2674 is predicted to result in loss of contacts in the hydrophobic core. (**m**) Interactions

822 involving Tyr2688 and Ile2694 in the core of the SET domain. The hydrogen bond
823 between Tyr2688 and Ser2661 is highlighted.
824

825 **Figure 5:**

826 ***KMT2B* Expression and Effects on Histone H3K4 Methylation**

827 (**a**) PCR analysis of human fetal and adult cDNA for expression of *KMT2B* (cropped gel
828 image; for uncropped image see **Supplementary Fig. 6**). *KMT2B* is widely expressed in
829 human tissues, including fibroblasts, brain tissue and midbrain dopaminergic neurons. (**b**)
830 Box plots of *KMT2B* mRNA expression levels in 10 adult brain regions (source:
831 BRAINEAC; http://www.braineac.org/). Expression levels are based on exon array
832 experiments as previously described and plotted on a log2 scale (y axis)[40]. *KMT2B* is
833 ubiquitously expressed across all 10 brain regions analyzed, with expression highest in the
834 cerebellum. Putamen (PUTM), frontal cortex (FCTX), temporal cortex (TCTX), occipital
835 cortex (OCTX). hippocampus (HIPP), substantia nigra (SNIG), medulla (specifically inferior
836 olivary nucleus, MEDU), intralobular white matter (WHMT), thalamus (THAL), and
837 cerebellar cortex (CRBL). “N” indicates the number of brain samples analyzed to generate
838 the results for each brain region. Whiskers extend from the box to 1.53 the interquartile
839 range. (**c**) Quantitative RT-PCR indicates that patients with *KMT2B* mutations (n = 4) have
840 significantly decreased fibroblast mRNA levels of *KMT2B* when compared to controls (n =
841 2) (Controls = 1.01±0.16SD; Patients = 0.57±0.12SD). n = 3 technical replicates were
842 analyzed per sample. Data were analyzed by two-tailed unpaired t-test: *P = 0.0182 (t =
843 3.856, df = 4). No significant difference in variances between the groups was detected by
844 F-test. (**d**) Histone methylation was assayed independently in three samples (n = 3;
845 technical replicates) taken from each patient-derived fibroblast cell line (n = 2; Patient 14
846 and 16) on different days, and compared with control cell lines (n = 2). Methylation values
847 are normalized to pan-histone H3 levels. Individual data-points are plotted with center bar

848 showing mean and error bars showing standard deviation. Differences between control
849 and patient-derived samples are not significant (H3K4me3 (left): Controls =
850 96.63±19.98SD; Patient 16 = 104.1±40.31SD; Patient 14 = 94.75±38.36SD; H3K4me2
851 (right): Controls = 94.33±19.25SD; Patient 16 = 127.8±20.79SD; Patient 14 =
852 80.23±31.09SD). Data were analyzed by one-way ANOVA: H3K4me3: P = 0.9196 (F =
853 0.08462, DFn = 2, DFd = 9); H3K4me2: P = 0.0727 (F = 3.557, DFn = 2, DFd = 9). (**e**)
854 Quantification of immunoblotting of tri-methyl H3K4 in *Dictyostelium* cell lysates. Tri-methyl
855 H3K4 intensity values are normalized against levels of total histone H3. H3K4 tri-
856 methylation is impaired in set1$^-$ cells compared to wild type. Expression of GFP-DdSet1 or
857 GFP-DdSet1(I1447T) in set1$^-$ cells rescues levels of H3K4Me3. Three independent point-
858 mutant cell lines (GFP-DdSet1(I1447T) 1-3) were created using the same point-mutant
859 DNA construct. Individual data-points (three independently prepared samples taken from
860 each cell line; n= 3, technical replicates) are plotted with center bar showing mean and
861 error bars showing standard deviation (Wild type = 115±48.25SD; set1$^-$ = 5.94±9.37SD;
862 set1$^-$ GFP-DdSet1(I1447T) 1 = 133.7±38.11SD; set1$^-$ GFP-DdSet1(I1447T) 2 =
863 129.8±42.34SD; set1$^-$ GFP-DdSet1(I1447T) 3 = 96.07±31.82SD; set1$^-$ GFP-DdSet1 =
864 110.5±12.02SD). No statistical testing was applied.

365 **Table 1: *KMT2B* Variants and Evolution of Motor Phenotype in KMT2B-dystonia**

Pat	Age (y) Sex M/F	*KMT2B* variants[(a)]	Symptoms at presentation: Body distribution & motor features	Onset of symptoms (y)	Bilateral LL involvement (y)	Bilateral UL involvement (y)	Onset of cranial, cervical, laryngeal dystonia (y)	Symptoms of cranial, cervical, laryngeal dystonia	Trial of medication and clinical response	Deep brain stimulation (DBS)
1	14 M	Deletion: Chr19: 35,608,666-36,233,508	RLL Right foot posturing Gait disturbance	4	6	6-11	5	Dysarthria Dysphonia Swallowing difficulties	L-dopa trial – no benefit	No
2	14 F	Deletion: Chr19: 35,197,252-38,140,100	Bilateral LL Limping Gait disturbance	7	7	8-11	8	Dysarthria Dysphonia Drooling	L-dopa trial – no benefit BLF – no benefit	No
3	9 M	Deletion: Chr19: 34,697,740-37,084,510	RLL Right foot posturing Gait disturbance	2.5	3	6-7	4	Dysarthria Dysphonia Swallowing difficulties Drooling	GBP – some reduction in tone	No
4	11 F	Deletion: Chr19: 36,191,100-36,376,860	LLL Left toe walking Gait disturbance	4	8	9-12	5	Dysarthria Dysphonia Swallowing difficulties Drooling	L-dopa trial– minimal benefit THP – minimal benefit	Planned for 2016
5	20 M	Deletion: Chr19: 31,725,360-36,229,548	Developmental delay Gait disturbance	Present but age of onset not known	Present but age of onset not known	Present but age of onset not known	Not known	Nasal voice	None	No
6	10 F	Deletion: Chr19: 35,017,972-36,307,788	RLL Right foot inversion	2.5	4	4	4-7	Dysarthria/anarthria Jaw-opening dystonia Swallowing difficulties NGF 6y PEG 8y Torticollis Severe retrocollis	L-dopa trial – no benefit THP – no benefit	Inserted age 7y Sustained excellent clinical benefits 3y post-DBS, marked improvement in torticollis, retrocollis, manual abilities and left leg dystonia. Loss of efficacy when 'DBS off' for almost a year and functional recovery when switched on again.
7	21 M	Deletion: Chr19: 35,414,997-37,579,142	RLL Right foot dragging Gait disturbance	7	7-8	13	13	Dysarthria Dysphonia Swallowing difficulties	L-dopa trial – no benefit BLF – no benefit	No

366

Pat	Age (y) Sex M/F	*KMT2B* variants [a]	Symptoms at presentation: Body distribution & motor features	Onset of symptoms (y)	Bilateral LL involve-ment (y)	Bilateral UL involve-ment (y)	Onset of cranial, cervical, laryngeal dystonia (y)	Symptoms of cranial, cervical, laryngeal dystonia	Trial of medication and clinical response	Deep brain stimulation (DBS)
8	17 F	Deletion: Chr19: 35,414,997-37,579,142	RLL Right foot posturing	4	6	4-12	2.5	Dysarthria Dysphonia Drooling Torticollis	L-dopa trial – no benefit	Inserted age 10y Good response over 6 years, particularly evident after replacement of faulty right DBS lead
9	14 M	Deletion: Chr19: 35,967,904-37,928,373	Bilateral LL Gait disturbance	4	4	9-13	9	Dysarthria Dysphonia	L-dopa trial – possible initial benefit but not sustained	Inserted age 14y Very good clinical response at 4m post-DBS with restoration of independent ambulation
10	7 F	Deletion: Chr19: 35,794,775-38,765,822	Bilateral LL Intermittent toe walking Gait disturbance	4	4	-	-	-	None	No
11	25 F	c.402dup p.Ser135Glnfs*23	RUL Right hand cramps and posturing	6	12	12	14[b]	Anarthria Orolingual dystonia Tongue thrusting Swallowing difficulties PEG	L-dopa trial – poorly tolerated, no benefit	Being considered
12	6 F	c.1690C>T p.Arg564*	Bilateral LL Toe walking	4	5	6	5	Dysarthria Swallowing difficulties	L-dopa trial – no benefit	No
13	11 M	c.3026_3027del p.Glu1009Glyfs*9	Bilateral UL Posturing, tremor Difficulty handwriting	8	9-10	8	9	Dysarthria Dysphonia	L-dopa trial – no benefit	No
14	18 M	c.3143_3149del p.Gly1048Glufs*132	Bilateral UL Posturing of hands Myoclonic jerks	8	13	8	13	Dysarthria Dysphonia Swallowing difficulties	L-dopa trial – no benefit	No
15	20 F	c.4545C>A p.Tyr1515*	Bilateral LL Toe Walking Clumsy	2	9	9	8.5	Dysarthria Dysphonia Oromandibular dystonia Swallowing difficulties PEG 18y	Moderate responses to (and currently taking) THP CLZ L-dopa BLF	No

367

Pat	Age (y) Sex M/F	*KMT2B* variants [(a)]	Symptoms at presentation: Body distribution & motor features	Onset of symptoms (y)	Bilateral LL involve-ment (y)	Bilateral UL involve-ment (y)	Onset of cranial, cervical, laryngeal dystonia (y)	Symptoms of cranial, cervical, laryngeal dystonia	Trial of medication and clinical response	Deep brain stimulation (DBS)
16	6 F	c.4688del p.Ala1563Aspfs*83	Bilateral LL Increasing falls Gait disturbance	3	3	5	6	Dysarthria Dysphonia	L-dopa trial – no benefit THP – initial benefit, not sustained	No
17	17 M	c.6515_6518delinsCCCAA p.Val2172Alafs*11	Bilateral LL Toe walking Gait disturbance	1	1	8	12	Dysarthria Dysphonia Swallowing difficulties	L-dopa trial – no benefit TBZ – no benefit BLF and THP – mild benefit	Inserted age 16y Very good clinical response 4m post-DBS with restoration of independent ambulation
18	20 F	c.8061del p.Tyr2688Thrfs*50	Clumsy movements Difficulties with speech articulation	1	-	-	Infancy	Dysarthria Dysphonia Swallowing and chewing difficulties	No	No
19	28 M	c.8079del p.Ile2694Serfs*44	Bilateral LL Toe walking Severe speech delay	2	3	4 (L>R)	7	Anarthria Jaw opening dystonia Tongue protrusion Swallowing difficulties PEG 8y L torticollis,R laterocollis	L-dopa trial – no benefit THP and TBZ reduced tongue protrusion	Inserted age 27y Improvement of jaw opening dystonia and tongue protrusion
20	40 M	c.3528+2T>A	LLL Gait disturbance L foot dragging Clumsiness	4	5	8	10	Severe dysarthria Dysphonia L torticollis	L-dopa trial – no benefit TBZ, THP, SUL – no benefit	Inserted age 32y – no benefit. Electrode replaced in 2009 with sustained improvement in foot posture but only transient benefit to cervical, UL and LL dystonia
21	18 M	c.4955G>A p.Gly1652Asp	RLL Right leg posturing	6	8	12	5	Dysarthria Dysphonia Swallowing difficulties	L-dopa trial – no benefit THP – not tolerated	Inserted age 15y Sustained clinical benefit 3y post-DBS, improved dystonia and independent walking
22	20 F	c.4986C>A p.Phe1662Leu	RLL Right foot posturing Abnormal gait	5	8	5-13	5-6	Dysarthria Dysphonia Swallowing difficulties Torticollis	L-dopa trial – no benefit BLF – no benefit THP – low dose, mild benefit BTX neck – reduction in pain, no functional benefit	Inserted age 20y Very good clinical response 9m post- DBS with improved dystonia and independent walking

368

Pat	Age (y) Sex M/F	*KMT2B* variants [a]	Symptoms at presentation: Body distribution & motor features	Onset of symptoms (y)	Bilateral LL involve-ment (y)	Bilateral UL involve-ment (y)	Onset of cranial, cervical, laryngeal dystonia (y)	Symptoms of cranial, cervical, laryngeal dystonia	Trial of medication and clinical response	Deep brain stimulation (DBS)
23	8 M	c.5114G>A p.Arg1705Gln	Bilateral LL Toe-walking	3	3	6	6.5	Dysarthria Torticollis	L-dopa trial – no benefit CLZ, THP, IT BLF – some benefit	Inserted age 7y with considerable benefit
24	27 F	c.5284C>T p.Arg1762Cys	LLL Tiptoe walking and in-turning of L foot	6	6	7	7	Dysarthria Anarthria from 14-15y Reduced tongue movements Swallowing preserved	L-dopa trial – no benefit THP- no benefit	No
25	19 F	c.5342T>C p.Leu1781Pro	RLL Right foot posturing Gait disturbance	8	12	13	10	Dysarthria Dysphonia Swallowing difficulties Torticollis	L-dopa trial – no benefit LVT – mild benefit	Inserted age 19y Very good clinical response 4m post-DBS with improved dystonia and ambulation[c]
26a	8 M	c.7549C>T p.Arg2517Trp	Delayed speech Delayed motor development	8	-	-	8	Severe paroxysmal retrocollis and jaw dystonia	-	No
26b	46 F	c.7549C>T p.Arg2517Trp	Bilateral UL UL posturing Torticollis Inability to walk long distances and run	23	26	23	23	Dysphonia Torticollis	None	No
27	19 F	c.8021T>C p.Ile2674Thr	RUL Posturing, tremor Difficulty handwriting Myoclonic jerks	9	11-13	10	9-10	Dysphonia	L-dopa trial – no benefit THP – no benefit LVT – no benefit CBZ – initial benefit, not sustained CLZ – not tolerated	No

369 BLF: baclofen; BTX: botulinum toxin; CLZ: clonazepam; GBP: gabapentin; IT: intrathecal; L: left; LL: lower limbs; LLL: left lower limb; LVT: levetiracetam; m: months; NGF: nasogastric
370 feeding; Pat: patient; PEG: percutaneous endoscopic gastrostomy; R: right; RLL: right lower limb; RUL: right upper limb; SUL: sulpiride; UL: upper limbs; TBZ: tetrabenzine; THP:
371 trihexyphenidyl; y: years
372 [a] based on NCBI Reference Sequence: NM_014727.2
373 [b] onset shortly after being fitted with orthodontic braces
374 [c] had undergone 2 posterior cranial fossa explorations and palatal surgery before DBS

375 **Table 2: Additional Clinical Features in Patient with *KMT2B* Variants**

Patient	*KMT2B* variants	Number of genes in microdeletion	Intellectual disability	Dysmorphic features	Additional neurological features	Psychiatric features	Abnormal skin features	Other systemic manifestations
1	Deletion: Chr19: 35,608,666-36,233,508	38	Mild	Elongated face	Not reported	Not reported	Not reported	Not reported
2	Deletion: Chr19: 35,197,252-38,140,100	124	No	Elongated face Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported
3	Deletion: Chr19: 34,697,740-37,084,510	109	Moderate	Elongated face	Not reported	Not reported	Cutis aplasia[(a)]	Retinal dystrophy
4	Deletion: Chr19: 36,191,100-36,376,860	14	V mild - subtle memory problems	Elongated face Broad nasal bridge Bulbous nasal tip	Not reported	Prone to anxiety[(b)]	Not reported	Not reported
5	Deletion: Chr19: 31,725,360-36,229,548	110	Moderate	Sparse hair Blepharophimosis Absent eyelashes of lower eyelids Low set, posteriorly rotated ears Epicanthic folds Narrow nasal bridge, ridge and point Largely bifid tongue Micrognathia Teeth overcrowding Finger contractures 5th finger clinodactyly Toe over-riding Dysplastic toenails	Microcephaly	Not reported	Occipital cutis aplasia	Small echogenic kidneys with low GFR, required renal transplant at 17 years
6	Deletion: Chr19: 35,017,97-36,307,788	69	No	Not reported	Microcephaly	Not reported	Not reported	Not reported
7	Deletion: Chr19: 35,414,997-37,579,142	99	Mild	Elongated face	Absence seizures	Not reported	Not reported	Absent right testis
8	Deletion: Chr19: 35,414,997-37,579,142	99	Mild	5th finger clinodactyly	Not reported	Not reported	Ectodermal dysplasia	Not reported
9	Deletion: Chr19: 35,967,904-37,928,373	79	Mild	Elongated face	Strabismus	Not reported	Not reported	Cleft palate
10	Deletion: Chr19: 35,794,775-38,765,822	111	Moderate	Not reported	Strabismus	Not reported	Not reported	Short stature Bronchiectasis

376

Patient	*KMT2B* variants	Number of genes in microdeletion	Intellectual disability	Dysmorphic features	Additional neurological features	Psychiatric features	Abnormal skin features	Other systemic manifestations
11	c.402dup p.Ser135Glnfs*23	-	No	Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported
12	c.1690C>T p.Arg564*	-	Moderate	Elongated face Bulbous nasal tip, short nasal root, Hypertelorism, large mouth with full lower lip	Epilepsy	Not reported	Not reported	Not reported
13	c.3026_3027del p.Glu1009Glyfs*9	-	V mild - difficulties with attention	Elongated face	Not reported	Not reported	Not reported	Not reported
14	c.3143_3149del p.Gly1048Glufs*132	-	No	Elongated face Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported
15	c.4545C>A p.Tyr1515*	-	No	Elongated face Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported
16	c.4688del p.Ala1563Aspfs*83	-	No	Elongated face	Not reported	Not reported	Not reported	Not reported
17	c.6515_6518delinsCCCAA p.Val2172Alafs*11	-	No	Elongated face	Not reported	Not reported	Phimosos	Short stature
18	c.8061del p.Tyr2688Thrfs*50	-	Mild	Micrognathia Atrophic tongue Bulbous nasal tip 5th finger clinodacytly	Not reported	Not reported	Not reported	Not reported
19	c.8079del p.Ile2694Serfs*44	-	No	Short stature	Delay in saccade initiation and hypometric vertical saccades	ADHD[3] with no response to Ritalin	Not reported	Not reported
20	c.3528+2T>A	-	Moderate 6y- verbal IQ 74 Performance IQ 87 No cognitive decline	Not reported	Not reported	Not reported	Not reported	Not reported
21	c.4955G>A p.Gly1652Asp	-	Mild	Elongated face	Not reported	Not reported	Not reported	Short stature Hypertrichosis
22	c.4986C>A p.Phe1662Leu	-	No	Elongated face Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported

377

378

Patient	*KMT2B* variants	Number of genes in microdeletion	Intellectual disability	Dysmorphic features	Additional neurological features	Psychiatric features	Abnormal skin features	Other systemic manifestations
23	c.5114G>A p.Arg1705Gln	-	Mild-moderate 6y WISC-IV 50-60	Elongated face Bulbous nasal tip Broad philtrum, Upslanted eyes, epicanthus, low-set ears, periorbital fullness, gap between front teeth	Spasticity in lower limbs from 6y	Not reported	Ichtyotic skin lesions with criss-cross pattern under the feet and at knees, broad scarring after operation	Episodic vomiting
24	c.5284C>T p.Arg1762Cys	-	No	Short stature	Oculomotor apraxia with difficulty initiating saccades. Mild spasticity	No	Not reported	Not reported
25	c.5342T>C p.Leu1781Pro	-	No	Elongated face Bulbous nasal tip	Not reported	Not reported	Not reported	Not reported
26a	c.7549C>T p.Arg2517Trp	-	No	Bulbous nasal tip	None	ADHD[c] Currently on methyphenidate, oxazepam, risperidone	Not reported	Not reported
26b	c.7549C>T p.Arg2517Trp	-	No	Bulbous nasal tip	Idiopathic intracranial hypertension – on acetazolamide	None	Not reported	Not reported
27	c.8021T>C p.Ile2674Thr	-	V subtle mild learning difficulties	Bulbous nasal tip	Not reported	Anxiety Self-harm behavior Depression Obsessive-compulsive traits[d]	Not reported	Not reported

379
380 [a] Supplementary Figure 4c
381 [b] Identified on formal psychology review
382 [c] Diagnosed by psychiatrist and under regular psychiatry review
383 [d] Under regular review with psychiatrist (ICD-10-CM F06.30; ICD-10-CM F42)
384 ADHD: attention deficit hyperactivity disorder; GFR: glomerular filtration rate; V: very; y: years

885 # ONLINE METHODS

886 **(1) Case Ascertainment**

887 Case ascertainment is summarized in **Supplementary Table 1** and
888 **Supplementary Fig. 1**. At Great Ormond Street-Institute of Child Health (GOS-
889 ICH), we identified 34 patients referred to our center with undiagnosed dystonia
890 (**Supplementary Table 1**). All patients (median age 13.5 years), presented with
891 progressive dystonia, with disease onset in childhood. None had a clinical history or
892 neuroimaging compatible with acquired dystonia, nor blood, urine or CSF biomarker
893 evidence of an underlying neurometabolic disorder. We used established national
894 and international clinical genetic and pediatric neurology networks to identify further
895 patients with microdeletions similar to those detected in the GOS-ICH cohort
896 (**Supplementary Fig. 1**). We also collaborated with research groups undertaking
897 whole exome sequencing in patients with early-onset dystonia (**Supplementary**
898 **Fig. 1**).

899 **(2) Molecular Genetic Analysis**

900 Genomic DNA was extracted from peripheral lymphocytes by standard techniques.
901 Written informed consent was obtained from participants, and all studies approved
902 by local ethics committees: National Research Ethics Service (NRES), London
903 Bloomsbury REC:13/LO/0168, Cambridge South REC:10/H0305/83; Republic of
904 Ireland REC:GEN/284/12; Human Research Ethics Committee (HREC),
905 HREC:10/CHW/114, 10/CHW/45; National Human Genome Research Institute
906 Institutional Research Board 76-HG-0238; Universities of Essen-Duisburg and
907 Erlangen-Nürnberg ethics committees Ref.3769; Medical Review Ethics Committee
908 Region Arnhem-Nijmegen, Ref:2011/188; UCL ethics committee, UCLH 06/N076.

909 Research was performed in accordance with the Declaration of Helsinki. Additional
910 consent for publication of photographs and videos was provided.

911 **Chromosomal Microarray**

912 Patients were analyzed for copy number variants using chromosomal microarray by
913 standard diagnostic techniques (**Supplementary Table 2**). Data is presented as
914 minimum coordinates in GRCh37/hg19.

915 **Whole Exome and Genome Sequencing (WES/WGS)**

916 WES/WGS was undertaken using center-specific protocols (see below). Reads
917 were aligned to the reference genome GRCh37/hg19. Detailed variant analysis was
918 performed for single nucleotide variants (SNVs) and small insertion/deletions
919 (indels) that (i) passed standard local quality filters, (ii) were predicted to alter
920 protein sequence in conserved residues, (iii) were predicted deleterious by
921 bioinformatics tools (including SIFT, PolyPhen-2, LRT, Mutation Taster, CADD), (iv)
922 had an allele frequency <0.01 in 1000 Genomes[41], NHLBI GO Exome Sequencing
923 Project, UK10K[42], Exome Aggregation Consortium (ExAC) database[8] and internal
924 control exomes/genomes. Data analysis was initially undertaken for known disease-
925 causing genes prior to analysis for autosomal recessive and dominant inheritance
926 models.

927 NIHR BioResource-Rare Diseases (NIHRBR-RD) Study: WGS was undertaken
928 using the Illumina TruSeq DNA PCR-Free Sample Preparation kit (Illumina Inc.,
929 San Diego, CA, USA) on Illumina Hiseq 2500, generating minimum coverage of
930 15X for ~95% of the genome, and average coverage of ~30X. Reads were aligned
931 using Isaac aligner (version 01.14) (Illumina Inc, Great Chesterford, UK)[43]. SNVs
932 and indels were identified using Isaac variant caller (version 2.0.17).

933 Wellcome Trust UK10K Rare Diseases project: DNA samples were captured using
934 Agilent SureSelect Target Enrichment V5 (Agilent Technologies, Santa Clara, CA,
935 USA) pull-down array. WES was performed on Illumina HiSeq 2000 platform.
936 Reads were aligned using the Burrows-Wheeler Alignment tool. SNVs and indels
937 were identified with SAMtools[44,45]. Variants were identified for each sample using
938 the Genome Analysis Toolkit (GATK) Unified Genotyper[46] and annotated with vcf-
939 annotate[47] and Ensembl Variant Effect Predictor v73 (VEP)[48].
940 National Institutes of Health, Bethesda: Exome sequencing was completed using
941 the TruSeqV2 exome capture kit. Data was aligned and processed as previously
942 decribed[49-51].
943 Institute of Human Genetics, Erlangen: Exome sequencing was performed on a
944 HiSeq 2500 (Illumina) platform with 125 bp paired-end sequencing using
945 SureSelect v.5 capturing reagents (Agilent).
946 Radboud University Medical Center, Nijmegen: Exome sequencing was undertaken
947 using Agilent SureSelectXT Human All Exon 50 Mb Kit, with sequencing on SOLiD
948 5500XL, producing an average sequence depth of 91X and average coverage of at
949 least 20X for 89% of targets. For calling and annotation of variants, a custom in-
950 house diagnostic pipeline was deployed[52].
951 UCL-Institute of Neurology, London: Exome sequencing was performed using
952 Illumina's Nextera Rapid Capture. Indexed and pooled libraries were sequenced on
953 Illumina's HiSeq3000 (100bp, paired-end). Reads were aligned with Novoalign.
954 Duplicate read removal, format conversion, and indexing were performed with
955 Picard. GATK was used to recalibrate base quality scores, perform local
956 realignments around possible indels, and to call (HaplotypeCaller) and filter (VQSR)
957 variants[46]. Annotated variant files were generated using ANNOVAR[53].

958 Deciphering Developmental Disorders study: Exome sequencing of family triomes
959 was performed using Agilent SureSelect Exome bait design (Agilent Human All-
960 Exon V3 Plus with custom ELID C0338371 and Agilent Human All-Exon V5 Plus
961 with custom ELID C0338371) on a Illumina HiSeq at the Wellcome Trust Sanger
962 Institute as previously described[54,55]. Data is currently available on 4,295 triomes
963 which were interrogated via a DDD complementary research proposal (CAP#120).

964 **Sanger Sequencing for Variant Validation and Gene Screening**

965 Direct sequencing was undertaken to (i) screen the entire coding region of *KMT2B*
966 for 13 cases from the GOS-ICH cohort (**Supplementary Table 1**), (ii) confirm
967 variants identified on next generation sequencing and (iii) establish familial
968 segregation (**Supplementary Fig. 2**). Additionally, cDNA from fibroblasts and
969 patient derived dopaminergic neurons were sequenced for a common SNP in exon
970 30 (rs231591). Primer pairs for all 37 coding exons and exon/intron boundaries of
971 *KMT2B* (Ensembl ENSG00000272333, transcript ENST00000420124) were
972 designed with Primer3 (**Supplementary Table 9**)[56,57]. PCR conditions can be
973 provided on request. PCR products were cleaned up (MicroCLEAN, Web Scientific)
974 and sequenced using the Big Dye Terminator Cycle Sequencing System (Applied
975 Biosystems Inc.). Sequencing reactions were run on an ABI PRISM 3730 DNA
976 Analyzer (Applied Biosystems Inc.) and analyzed using Chromas.

977 **Enrichment Analysis**

978 The number of *de novo* predicted protein truncating variants (PPTVs) in *KMT2B*
979 expected to be seen by chance in a subset of the NIHR BioResource–Rare
980 Diseases cohort with pediatric onset neurological disease (n=272), was calculated
981 using published gene-specific mutation rates[58] and scaled to account for frameshift,
982 nonsense and essential splice site variants[58]. To assess significance, the expected

983 number of *de novo* PPTVs were compared to the observed number, assuming a
984 Poisson distribution.

985 **(3) CSF Neurotransmitter Analysis**

986 CSF was collected by lumbar puncture and diagnostically analyzed for
987 neurotransmitter monoamine metabolites in specialist laboratories (London,
988 Barcelona, Sydney, Jerusalem) by high performance liquid chromatography[59.60]

989 **(4) Comparative Modelling**

990 *In silico* homology modeling was utilized to predict putative effects of *KMT2B*
991 variants. The Pfam database[61] was used to assign known domains to the full-length
992 sequence of KMT2B. Evolutionary conservation of residues in the sequence was
993 quantified using Consurf server[62], based on alignment with a set of homologous
994 sequences, which share 35%-95% sequence identity with KMT2B. HHpred[63] was
995 utilized to identify proteins or domains with known structure that have similar
996 sequence and structural features to KMT2B. Selected templates had more than
997 99% probability (based on HHpred alignment score) of being related structurally to
998 specific domain segments of KMT2B. MODELLER[64] was employed for different
999 regions of KMT2B and HHpred alignments were used to dictate residue
1000 equivalences with the template. For each domain, 150 models were generated with
1001 MODELLER loop optimization protocol and the best model was selected based on
1002 the normalized DOPE score[65]. The effect of a point substitution on the stability of
1003 the domain structure was evaluated using DUET[11] and MAESTRO[10]. Visualization
1004 and analysis of amino acid interactions and generation of mutant models were done
1005 with UCSF Chimera[66].

1006 **(5) Histone Methylation Assays**

1007 **H3K4 Methylation**

1008 Histones were extracted from fibroblasts using a modified version of a published
1009 protocol[67]. Cells were lysed by rotating at 4°C in hypotonic lysis buffer (10mM Tris-
1010 Cl pH8.0, 1mM KCl, 1.5mM $MgCl_2$, 1mM DTT, Roche complete protease inhibitors).
1011 Intact nuclei were pelleted by centrifugation and resuspended in 0.2N HCl.
1012 Following overnight histone extraction by rotating at 4°C, nuclear debris was
1013 removed by centrifugation and soluble histones precipitated by dropwise addition of
1014 TCA to a final concentration of 33%. Following one hour precipitation on ice,
1015 histones were pelleted by centrifugation and washed with acetone before
1016 resuspension in Milli-Q water.

1017 **Expression of p.Ile1447Thr Set1 Point Substitution in *Dictyostelium***

1018 *Dictyostelium discoideum* cells were grown as previously described[68]. They are not
1019 listed in the database of commonly misidentified cell lines maintained by ICLAC.
1020 *Dictyostelium* strains[69] included wild type—AX2 (DBS0238015) and *set1*$^-$KO—*set1*$^-$
1021 (DBS0236928). DdSet1 has Dictybase gene ID DDB_G0289257 and Uniprot ID
1022 Q54HS3. p.Ile1447Thr was created by substituting ATT for ACT in a *DdSet1*
1023 genomic clone by PCR (**Supplementary Table 10**). The product containing this
1024 substitution was cloned as a ClaI/EcoRI fragment replacing the equivalent region of
1025 a wild type *DdSet1* genomic clone. This region was subsequently subcloned as a
1026 ClaI/AccI fragment into a pDEXH[70] based integrating plasmid containing *GFP-*
1027 *DdSet1* under control of the *DdAct15* promoter and a G418 selection marker –
1028 replacing the same region of the wild type *DdSet1* cDNA sequence. The presence
1029 of p.Ile1447Thr in the resulting plasmid, pJN106, was confirmed by Sanger
1030 sequencing. Constructs for expression of GFP-DdSet1(p.Ile1447Thr) (pJN106) and

1031 wild type GFP-DdSet1 (pJRC18) were transformed into *set1⁻ Dictyostelium* cells[71]
1032 as previously described[68]. Transformants were selected by addition of 10ug/ml
1033 Geneticin (Gibco) to growth medium. Expression of full-length wild type and point
1034 mutant GFP-DdSet1 was confirmed by anti-GFP immunoblotting.

1035 **Immunoblot Analysis of Histone Methylation**

1036 Fibroblast histone samples were diluted in SDS sample buffer (Bio-Rad) containing
1037 5% [v/v] β-mercaptoethanol and protease inhibitors (Roche Complete), separated
1038 by SDS-PAGE, then blotted onto nitrocellulose. Histone H3 and methylated K4
1039 variants were detected using rabbit polyclonal anti-histone H3 (Abcam ab1791,
1040 1:1000 dilution), rabbit polyclonal anti-histone H3 tri-methyl K4 (Abcam ab8580,
1041 1:1000 dilution), rabbit polyclonal anti-histone H3 di-methyl K4 (Millipore 07-030,
1042 1:2000 dilution). Secondary antibody used was donkey anti-rabbit IgG HRP-
1043 conjugated (GE Healthcare NA934V; for Histone H3 and tri-methyl K4 detection
1044 1:30000 dilution, for di-methyl K4 detection 1:20000 dilution). Following detection
1045 using Supersignal West Pico chemiluminescent substrate (Thermo) and CL-
1046 Xposure film (Thermo), densitometry was performed using ImageJ[72].

1047 *Dictyostelium* cells were collected by centrifugation and resuspended in KK_2
1048 buffer[68] before lysis in SDS sample buffer (Bio-Rad) containing 5% [v/v] β-
1049 mercaptoethanol and protease inhibitors (Roche Complete). Immunoblotting for
1050 GFP-DdSet1 expression was assayed as above, using a mouse IgG monoclonal
1051 anti-GFP primary (Roche 11814460001, 1:500 dilution) and anti-mouse IgG HRP-
1052 conjugated secondary antibody (BioRad 170-6516, 1:20000 dilution).
1053 Immunoblotting for Histone H3 and tri-methyl H3K4 was conducted as for
1054 fibroblasts (with the modification: anti-histone H3 tri-methyl K4 dilution 1:3000).

1055 **(6) RNA and Protein Measurements**

1056 **Fibroblast RNA Extraction and cDNA Synthesis**

1057 Skin biopsies from Patients 2, 13, 14 and 16 were taken for fibroblast culture, and
1058 grown in Dulbecco's Modified Eagle's Medium (DMEM, Sigma) with 4.5g/L glucose,
1059 4mM L-glutamine, and 10% heat inactivated fetal bovine serum (Life Technologies)
1060 and maintained in an incubator at 37°C and 5% CO_2. Fibroblasts from two age-
1061 matched controls were supplied by the Dubowitz Neuromuscular Centre Biobank
1062 (GOS-ICH). Cultures were checked for mycoplasma contamination (MycoAlert
1063 Mycoplasma Detection Kit, Lonza). As fibroblast cultures were derived from human
1064 skin biopsies, no authentication was undertaken. Furthermore these cell lines do
1065 not belong to the commonly misidentified cell lines listed in the database
1066 maintained by ICLAC. RNA was extracted from fibroblasts of T75 cell culture flasks
1067 using the RNeasy Mini Kit from QIAGEN. First-Strand cDNA synthesis was carried
1068 out with SuperScript® III Reverse Transcriptase (Invitrogen) using 500ng total RNA
1069 per reaction and Oligo (dT) primers (Thermo Fisher Scientific).

1070 **Generation of Dopaminergic Neurons, RNA Extraction, cDNA Synthesis**

1071 Fibroblasts from a KMT2B-negative individual were reprogrammed into induced
1072 pluripotent stem cells (iPSC) using an established Sendai virus protocol
1073 (CytoTune®-iPS Reprogramming Kit, Invitrogen)[73]. iPSC lines were stringently
1074 tested for pluripotency using established methods[74] before differentiation into
1075 dopaminergic neurons[75]. After 60 days of differentiation, dopaminergic identity was
1076 confirmed by immunofluorescence for neuronal marker, MAP2, (mouse monoclonal
1077 anti-MAP2, Sigma, M9942, 1:400 dilution) and dopaminergic marker, TH (chicken
1078 polyclonal anti-TH, Aves labs, TYH, 1;400 dilution). Nuclei were contrasted with
1079 DAPI. Microscopic images were captured (Zeiss LSM710 Confocal) and analyzed

1080 using ImageJ[72]. Neuronal differentiation efficiency was determined by calculating
1081 the number of MAP2/TH positive cells relative to MAP2-positive cells
1082 (**Supplementary Fig. 14**). RNA extraction and cDNA synthesis was carried out as
1083 described for fibroblasts.

1084 **PCR Analysis**

1085 We investigated tissue expression of *KMT2B* in (i) human fetal cDNA samples
1086 (Moore fetal tissue cohort)[76], (ii) a human cDNA panel (Clontech), (iii) human
1087 fibroblasts and (iv) dopaminergic neurons differentiated from human iPSC. PCR
1088 amplification of cDNA (**Supplementary Table 11**) was performed with BioMix™
1089 Red (Bioline Ltd, conditions available on request). PCR products were separated on
1090 a 2% agarose gel containing Ethidium bromide (Sigma) and visualized with Gel
1091 Doc™ XR+ System (Bio Rad).

1092 Changes in relative expression of *KMT2B*, *THAP1* and *TOR1A*, were measured by
1093 quantitative RT-PCR on a StepOnePlus™ Real-Time PCR System (Applied
1094 Biosystems). RT-PCR reactions comprised 1x MESA Blue qPCR MasterMix Plus
1095 for SYBR® Assay (Eurogentec), 0.1µl ROX Reference Dye (Invitrogen), 9µL cDNA
1096 (of a dilution 1:25) and 500nM of each primer (**Supplementary Table 11**). RT-PCR
1097 conditions are available on request. Relative quantification of gene expression was
1098 determined using the $2^{-\Delta\Delta Ct}$ method[77], with glyceraldehyde-3-phosphate
1099 dehydrogenase (*GAPDH*) as a reference gene.

1100 **Fibroblast Protein Preparation and Immunoblot Analysis**

1101 Fibroblasts grown in T25 cell culture flasks were washed with cold PBS and
1102 incubated with lysis buffer [150mM NaCl, 50mM Tris pH8, 1% NP40 and 1x
1103 cOmplete™ Mini Protease Inhibitor Cocktail (Roche)] for 30 minutes on ice. Lysed
1104 cells were centrifuged at 13,000 rpm for 15 minutes to remove cell debris. Protein

1105 concentrations of the cell lysates were measured with the Pierce™ BCA Protein
1106 Assay Kit (Thermo Fisher Scientific). A total of 5-10ng protein were prepared with
1107 1x Laemmli buffer and 0.5M DTT and boiled for 5 minutes at 100°C for denaturing.
1108 Proteins were separated by electrophoresis on 4–20% Mini-PROTEAN® TGX
1109 Stain-Free™ Protein Gels (Bio Rad) by applying 300V for ~17 minutes. Proteins
1110 were transferred to polyvinylidene difluoride (PVDF) membranes (Bio Rad) using
1111 the Trans-Blot® Turbo™ Transfer System (Bio Rad). Membranes were incubated
1112 for 1 hour at room temperature in blocking solution (5% nonfat dry milk in
1113 Phosphate-buffered saline-Tween 20, PBS-T) and then probed with polyclonal
1114 rabbit anti-THAP1 (Cambridge Bioscience [Supplier: Proteintech], 12584-1-AP,
1115 1:1500 dilution) and monoclonal mouse anti-TorsinA (Cell Signaling, D-M2A8,
1116 1:1000 dilution), respectively, in blocking buffer (1% nonfat dry milk in PBS-T;
1117 except THAP1 antibody which was diluted in 5% nonfat dry milk in PBS-T) for
1118 approximately 16 hours at 4°C. Following three washing steps with PBS-T,
1119 membranes were incubated for 1 hour at room temperature with horseradish
1120 peroxidase (HRP)-conjugated goat anti-rabbit IgG antibody (Cell Signaling, #7074,
1121 1:3000 dilution) and HRP-conjugated horse anti-mouse IgG antibody (Cell
1122 Signaling, #7076, 1:3000 dilution), respectively. Afterwards the blot was washed
1123 three times with PBS-T and signals were visualized with ClarityTM ECL Western
1124 Blotting Substrate (Bio Rad) on a Gel Doc™ XR+ System (Bio Rad). To confirm
1125 equivalent loading, blots were stripped at 37°C for 15 minutes in Restore™ Western
1126 blot Stripping buffer (Thermo Fisher Scientific), blocked for 1 hour, and reprobed
1127 with HRP-conjugated rabbit anti-β-Tubulin (Cell Signaling, 9F3, 1:1000 dilution). For
1128 quantification, intensity values of control and patient bands were determined using

1129 Fiji software[78] and normalized against the intensity value of the reference protein
1130 band.

1131 **CSF Immunoblotting**

1132 CSF protein levels of tyrosine hydroxylase (TH) and dopamine receptor D2 (D2R)
1133 were analyzed. CSF samples were available from two patients (Patients 2 and 16)
1134 and four gender and age-matched controls (patients with no history of movement
1135 disorder, on no medication). Immunoblotting was carried out as described above.
1136 For the detection of TH and D2R the membranes were incubated with polyclonal
1137 rabbit anti-TH (Millipore, AB152, 1:1000 dilution) and polyclonal rabbit anti-D2R
1138 (Millipore, AB5084P, 1:1000 dilution), respectively, followed by 2 hours incubation
1139 with HRP-conjugated goat anti-rabbit IgG antibody (Cell Signaling, #7074, 1:3000
1140 dilution). As an internal control for loading monoclonal mouse anti-Transferrin
1141 (Santa Cruz, E-8, 1:1000 dilution) followed by HRP-conjugated horse anti-mouse
1142 IgG antibody (Cell Signaling, #7076, 1:3000 dilution) were used.

1143 **(7) Statistics**

1144 The statistical analyses for the histone methylation assays were conducted using
1145 GraphPad Prism v7.01 and for the analyses of the fibroblast cell lines and CSF
1146 immunoblotting using GraphPad v5. The final data are represented with the mean
1147 and the standard deviation as error bars. For multiple comparisons one-way
1148 ANOVA was performed whereas for dual comparisons unpaired two-tailed
1149 Student's t test were employed. $P < 0.05$ was considered significant: *$P < 0.05$, **P
1150 < 0.01, ***$P < 0.001$. The F test was utilized to compare the variances between the
1151 groups in dual comparisons.

1152 We assume that technical replicates of immunoblot assays using the same cell
1153 lines will be normally distributed. For the fibroblast histone methylation assay the

1154 Brown-Forsythe test was used to check differences in variance between the groups
1155 compared, and no significant differences was found in standard deviation
1156 (H3K4Me3: $p = 0.7567$ [$F = 0.2877$, DFn = 2, DFd = 9]; H3K4me2: $p=0.8446$ [$F =$
1157 0.1721, DFn = 2, DFd = 9]). For the remaining experiments data distribution was
1158 not tested but was assumed to be normal. Blinding was not applied for data
1159 collection and analysis.

1160 ## References

1161 41. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation
1162 from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
1163 42. UK10K Consortium *et al.* The UK10K project identifies rare variants in health and
1164 disease. *Nature* **526**, 82-90 (2015).
1165 43. Raczy, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina
1166 sequencing platforms. *Bioinformatics* **29**, 2041-2043 (2013).
1167 44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1168 transform. *Bioinformatics* **25**, 1754-1760 (2009).
1169 45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
1170 2078-2079 (2009).
1171 46. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
1172 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303
1173 (2010).
1174 47. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-
1175 2158 (2011).
1176 48. McLaren, W. *et al.* Deriving the consequences of genomic variants with the
1177 Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-2070 (2010).
1178 49. Gahl, W.A. *et al.* The National Institutes of Health Undiagnosed Diseases Program:
1179 insights into rare diseases. *Genet. Med.* **14**, 51-59 (2012).
1180 50. Adams, D.R. *et al.* Analysis of DNA sequence variants detected by high-throughput
1181 sequencing. *Hum. Mutat.* **33**, 599-608 (2012).
1182 51. Bone, W.P. *et al.* Computational evaluation of exome sequence data using human
1183 and model organism phenotypes improves diagnostic efficiency. *Genet. Med.* **18**,
1184 608-617 (2016).

1185 52. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual
1186 disability. *N. Engl. J. Med.* **367**, 1921-1929 (2012).
1187 53. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
1188 variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164
1189 (2010).
1190 54. Akawi, N. *et al.* Discovery of four recessive developmental disorders using
1191 probabilistic genotype and phenotype matching among 4,125 families. *Nat. Genet.*
1192 **47**, 1363-1369 (2015).
1193 55. Deciphering Developmental Disorders Study. Large-scale discovery of novel
1194 genetic causes of developmental disorders. *Nature* **519**, 223-228 (2015).
1195 56. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.*
1196 **40**, e115 (2012).
1197 57. Koressaar, T. & Remm, M. Enhancements and modifications of primer design
1198 program Primer3. *Bioinformatics* **23**, 1289-1291 (2007).
1199 58. Samocha, K.E. *et al.* A framework for the interpretation of de novo mutation in
1200 human disease. *Nat. Genet.* **46**, 944-950 (2014).
1201 59. Hyland, K. *et al.* Cerebrospinal fluid concentrations of pterins and metabolites of
1202 serotonin and dopamine in a pediatric reference population. *Pediatr. Res.* **34**, 10-14
1203 (1993).
1204 60. Ormazabal, A. *et al.* HPLC with electrochemical and fluorescence detection
1205 procedures for the diagnosis of inborn errors of biogenic amines and pterins. *J.*
1206 *Neurosci. Methods* **142**, 153-158 (2005).
1207 61. Finn, R.D. *et al.* The Pfam protein families database: towards a more sustainable
1208 future. *Nucleic Acids Res.* **44**, D279-285 (2016).
1209 62. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores
1210 of residues on protein structures. *Nucleic Acids Res.* **33**, W299-302 (2005).

1211 63. Söding, J., Biegert, A. & Lupas, A.N. The HHpred interactive server for protein
1212 homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244-248
1213 (2005).
1214 64. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial
1215 restraints. *J. Mol. Biol.* **234**, 779-815 (1993).
1216 65. Shen, M.Y. & Sali, A. Statistical potential for assessment and prediction of protein
1217 structures. *Protein Sci.* **15**, 2507-2524 (2006).
1218 66. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory
1219 research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
1220 67. Shechter, D., Dormann, H.L., Allis, C.D. & Hake, S.B. Extraction, purification and
1221 analysis of histones. *Nat. Protoc.* **2**, 1445-1457 (2007).
1222 68. Fey, P., Kowal, A.S., Gaudet, P., Pilcher, K.E. & Chisholm, R.L. Protocols for
1223 growth and development of Dictyostelium discoideum. *Nat. Protoc.* **2**, 1307-1316
1224 (2007).
1225 69. Fey, P., Dodson, R.J., Basu, S. & Chisholm, R.L. One stop shop for everything
1226 Dictyostelium: dictyBase and the Dicty Stock Center in 2012. *Methods Mol. Biol.*
1227 **983**, 59-92 (2013).
1228 70. Faix, J., Gerisch, G. & Noegel, A.A. Overexpression of the csA cell adhesion
1229 molecule under its own cAMP-regulated promoter impairs morphogenesis in
1230 Dictyostelium. *J. Cell Sci.* **102**, 203-214 (1992).
1231 71. Chubb, J.R. *et al.* Developmental timing in Dictyostelium is regulated by the Set1
1232 histone methyltransferase. *Dev. Biol.* **292**, 519-532 (2006).
1233 72. Rasband, W.S. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland,
1234 USA, http://imagej.nih.gov/ij/, (1997-2015).
1235 73. Fusaki, N., Ban, H., Nishiyama, A., Saek,i K. & Hasegawa, M. Efficient induction of
1236 transgene-free human pluripotent stem cells using a vector based on Sendai virus,

1237 an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad. Ser. B*
1238 *Phys. Biol. Sci*. **85**, 348-362 (2009).
1239 74. Hartfield, E.M. *et al.* Physiological characterisation of human iPS-derived
1240 dopaminergic neurons. *PLoS One* **9**, e87388 (2014).
1241 75. Kirkeby, A. *et al.* Generation of regionally specified neural progenitors and
1242 functional neurons from human embryonic stem cells under defined conditions. *Cell*
1243 *Rep.* **1**, 703-714 (2012).
1244 76. Frost, J.M. *et al.* Evaluation of allelic expression of imprinted genes in adult human
1245 blood. *PLoS One* **5**, :e13556 (2010).
1246 77. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-
1247 time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408
1248 (2001).
1249 78. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat.*
1250 *Methods* **9**, 676-682 (2012).