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

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

Affiliations:

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

2. Department of Haematology, University of Cambridge, NHS Blood and Transplant Centre, Cambridge, UK.
3. NIHR BioResource-Rare Diseases, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, UK.
4. Clinical Genetics, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK.
5. MRC Laboratory for Molecular Cell Biology and Department of Cell and Developmental Biology, UCL, London, UK.
6. Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK.
7. Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.
8. Institute of Structural and Molecular Biology, Crystallography/Department of Biological Sciences, Birkbeck College, University of London, London, UK.
9. Department of Molecular Neuroscience, UCL-Institute of Neurology, London, UK.
10. Department of Neurology, Great Ormond Street Hospital, London, UK.
11. Pediatric Neurology and Development, Shaare-Zedek Hospital, Jerusalem, Israel.
12. Department of Paediatric Neurology, Donders Centre for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, Netherlands.
13. Department of Neurology, Hadassah Medical Center and the Hebrew University, Jerusalem, Israel.
14. Department of Clinical Genetics, Great Ormond Street Hospital, London, UK.
15. Department of Neurobiology and Neurosurgery, The Hebrew University, Hadassah Medical Centre, Jerusalem, Israel.

16. Victoria Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Victoria, Australia.
17. Department of Pediatrics, Institute of Clinical Sciences, University of Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden.
18. Department of Clinical Genetics, Southampton General Hospital, Southampton, UK.
19. Department of Neurology, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK.
20. Bristol Genetics Laboratory, Bristol, UK.
21. Functional and Restorative Neurosurgery, Hadassah University Hospital, Jerusalem, Israel.
22. Complex Motor Disorders Service, Evelina Children's Hospital, Guy's & St Thomas' NHS Foundation Trust, London, UK.
23. Sobell Department, National Hospital for Neurology and Neurosurgery, London, UK.
24. Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast, UK.
25. Child and Adolescent Health, University of Sydney, Sydney, Australia.
26. Institute for Neuroscience and Muscle Research, The Children's Hospital at Westmead, University of Sydney, Sydney, Australia.
27. Department of Neurology, Maastricht University Medical Center, Maastricht, Netherlands.
28. Department of Pediatrics, Piteå Hospital & Umeå University Hospital, Umeå, Sweden.

29. College of Medicine and Dental Studies, The University of Birmingham, Birmingham, UK.
30. Neuroscience and Mental Health Research Institute, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK.
31. Department of Cytogenetics, The Children's Hospital at Westmead, Westmead, Australia.
32. Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
33. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, Netherlands.
34. Medical Genetics Institute and Pediatrics, Shaare Zedek Medical Center and the Hebrew University School of Medicine, Jerusalem, Israel.
35. Department of Clinical Genetics and School for Oncology and Developmental Biology (GROW), Maastricht University Medical Center, Maastricht, Netherlands.
36. Department of Paediatric Neurology, John Radcliffe Hospital, Oxford, UK.
37. Department of Paediatrics, University of Melbourne, Melbourne, Australia.
38. Institute of Human Genetics, University Duisburg-Essen, Essen, Germany.
39. Institute of Human Genetics, Heinrich-Heine-University, Medical Faculty, Düsseldorf, Germany.
40. Neurometabolic Unit, National Hospital for Neurology and Neurosurgery, Queen Square, London, UK.
41. Clinical Chemistry, Great Ormond Street Hospital, NHS Foundation Trust, London, UK.

42. North East Thames Regional Genetics Service, Great Ormond Street Hospital, London, UK.
43. The members of this consortium/study are listed in the **Supplementary Note**.
44. Department of Child Neurology, Hospital Sant Joan de Déu, Universitat de Barcelona, Barcelona, Spain.
45. Centre for Biomedical Research in Rare Diseases (CIBERER-ISCIII), Hospital Sant Joan de Déu, Barcelona, Spain.
46. NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA.
47. Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands.
48. Department of Radiology, Great Ormond Street Hospital, London, UK.
49. These authors contributed equally to this work.
50. These authors contributed equally to this work.

Correspondence should be addressed to M.A.K. (manju.kurian@ucl.ac.uk)

ABSTRACT

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

INTRODUCTION

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

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

RESULTS

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

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

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

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

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

Phenotypic characterization of patients with KMT2B variants

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

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

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

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

Deep brain stimulation: clinical benefit in KMT2B-dystonia

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

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

KMT2B is constrained for missense and predicted protein truncating variants

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

PPTVs were observed. This represents a significant enrichment (p-value $3.12x10^{-08}$). 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].

KMT2B variants are predicted to destabilize protein structure

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**).

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

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

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

KMT2B is ubiquitously expressed with reduced expression in KMT2B-dystonia

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

Histone H3K4 methylation is not globally reduced in KMT2B-dystonia

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

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

Altered gene and protein expression in KMT2B-dystonia

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

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

DISCUSSION

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

For many patients, KMT2B-dystonia is associated with a number of additional clinical features including other neurological symptoms, intellectual disability, psychiatric co-morbidity, dysmorphia, skin lesions and other systemic signs. Given the association with active gene expression, it is conceivable that *KMT2B* variants 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)

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

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

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

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

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

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

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

URLs:

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

http://exac.broadinstitute.org

DECIPHER

http://decipher.sanger.ac.uk

UK10K Project

http://www.uk10k.org

Deciphering Developmental Disorders (DDD) study

http://www.ddduk.org/

1000 Genomes

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

NHLBI GO Exome Sequencing Project (release 20130513)

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

Ensembl genome browser

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

Primer3

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

Chromas Sequencing software

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

Clustal Omega

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

SIFT

http://sift.jcvi.org/

PolyPhen-2

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

Mutation Taster

http://www.mutationtaster.org/

Combined Annotation Dependent Depletion (CADD)

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

BRAINEAC

http://www.braineac.org.

Methods:

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

Accession codes:

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

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

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

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

Patient 13 (Ref: EGAR00001320121); Patient 14 (Ref: EGAR00001314777); Patient 17 (Ref: EGAR00001314751) and Patient 21 (Ref: EGAR00001314767).

https://www.ebi.ac.uk/ega/home

UK10K Project: UK10K whole exome sequencing data has been deposited under the name UK10K_RARE_FIND. Accession code: EGAS00001000128. Title of dataset: UK10K_RARE_FIND REL-2013-10-31 variant calling: EGAD00001000750 Data is deposited for Patients 22 (Ref: UK10K_FIND5536224) and 27 (Ref: UK10K_FIND5536279).

https://www.ebi.ac.uk/ega/studies/EGAS00001000128

Deciphering Developmental Disorders (DDD) study: Exome sequencing data is accessible via the European Genome-phenome Archive (EGA) under accession EGAS00001000775.

https://www.ebi.ac.uk/ega/studies/EGAS00001000775

National Institutes of Health, Bethesda; Institute of Human Genetics, Erlangen; Radboud University Medical Center, Nijmegen, UCL-Institute of Neurology, London: Whole exome sequencing data has not been deposited since participating patients have not consented for the data to be publicly released.

Note:

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

Acknowledgements:

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

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

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

Author contributions

E.M., K.J.C., J.M.E.N., J.R.C., F.L.R. and M.A.K. conceived and designed 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., 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.,

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

Competing financial interests

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

References

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

(MLL) protein via the histone H3-binding pocket. *J. Biol. Chem.* **283**, 35258-35264 (2008).

19. Chubb, J.R. *et al.* Developmental timing in Dictyostelium is regulated by the Set1 histone methyltransferase. *Dev. Biol.* **292**, 519-532 (2006).
20. Malan, V. *et al.* 19q13.11 deletion syndrome: a novel clinically recognisable genetic condition identified by array comparative genomic hybridisation. *J. Med. Genet.* **46**, 635-640 (2009).
21. Kruer, M.C. *et al.* Neuroimaging features of neurodegeneration with brain iron accumulation. *AJNR Am. J. Neuroradiol.* **33**, 407-414 (2012).
22. Meyer, E., Kurian, M.A. & Hayflick, S.J. Neurodegeneration with Brain Iron Accumulation: Genetic Diversity and Pathophysiological Mechanisms. *Annu. Rev. Genomics Hum. Genet.* **16**, 257-279 (2015).
23. Ozelius, L. *et al.* SourceGeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; [updated 2014 Jan 02] (1993-2016).
24. Klein, C. *et al.* SourceGeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; [updated 2014 May 1] (1993-2016).
25. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
26. Black, J.C., Van Rechem, C. & Whetstine, J.R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**, 491-507 (2012).
27. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U S A* **107**, 21931-21936 (2010).
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).

Figure legends

Figure 1:

Molecular Genetics Findings in Patients with *KMT2B* Variants

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

Figure 2:

Clinical Features of Patients with *KMT2B* Variants

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

Figure 3:

Radiological Features of Patients with KMT2B Variants

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

(**a,e,i**) Representative MRI from control subjects for T2*-weighted sequences (**a**: age 10y2m), diffusion-weighted sequences (**e**: age 10y4m) and susceptibility weighted sequences (**i**: age 10y8m) indicating normal appearances of basal ganglia. Patient 1, age 9y5m (**b,f,j**), Patient 13, age 11y3m (**c,g,k**), Patient 9, age 15y1m (**d**), Patient 22, age 13y1m (**h**) and Patient 25, age 16y (**l**): evidence of bilateral subtle hypointensity of the globus pallidus with hypointense lateral streak of globus pallidus externa.

Figure 4:

Comparative Modelling of KMT2B Protein Structure

(**a**) Schematic domain architecture of KMT2B. (**b-d**) Degree of amino conservation is displayed in the structural models for different domains. Red to blue indicates increasing conservation. (**b**) Model of PHD-like domain shows Gly1652 and Phe1662. (**c**) Model of FYRN domain presents position and conservation of Arg1762 and Leu1781. (**d**) Model of the SET methyltransferase domain indicates position and conservation of Ile2674, Tyr2688 and Ile2694. (**e**) Location of Gly1652 in the PHD-like domain model and the hydrogen bond network in the vicinity (α3 helix is indicated). (**f**) Hydrophobic packing involving Phe1662. (**g**) Change to leucine at 1662 is predicted to cause loss of contacts within the hydrophobic core. Residue side chains are presented as spheres highlighting van der Waals contacts. (**h**) Interactions involving Arg1762 from FYRN with Arg2463 and Leu2464 of FYRC. The hydrogen bond interactions and distances are highlighted. (**i**) Leu1781 shown at the interface of FYRN (orange)/FYRC (magenta) domains. The backbone hydrogen bonds stabilizing the sheet structure are highlighted. (**j**) Interactions involving Arg2517 and WDR5 (beige). The salt bridge interaction between Arg2517 of KMT2B and Asp172 of WDR5 is highlighted. (**k**) Location and contacts involving Ile2674 in the hydrophobic core of the SET domain (SAH is indicated). (**l**) Substitution with threonine at 2674 is predicted to result in loss of contacts in the hydrophobic core. (**m**) Interactions

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

Figure 5:

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

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

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

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 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)
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

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

867

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

868

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

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 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: trihexyphenidyl; y: years

[(a)] based on NCBI Reference Sequence: NM_014727.2

[(b)] onset shortly after being fitted with orthodontic braces

[(c)] had undergone 2 posterior cranial fossa explorations and palatal surgery before DBS

875 **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 clinodacytly	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

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

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

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

ONLINE METHODS

(1) Case Ascertainment

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

(2) Molecular Genetic Analysis

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

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

Chromosomal Microarray

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

Whole Exome and Genome Sequencing (WES/WGS)

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

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

Wellcome Trust UK10K Rare Diseases project: DNA samples were captured using Agilent SureSelect Target Enrichment V5 (Agilent Technologies, Santa Clara, CA, USA) pull-down array. WES was performed on Illumina HiSeq 2000 platform. Reads were aligned using the Burrows-Wheeler Alignment tool. SNVs and indels were identified with SAMtools[44,45]. Variants were identified for each sample using the Genome Analysis Toolkit (GATK) Unified Genotyper[46] and annotated with vcf-annotate[47] and Ensembl Variant Effect Predictor v73 (VEP)[48].

National Institutes of Health, Bethesda: Exome sequencing was completed using the TruSeqV2 exome capture kit. Data was aligned and processed as previously decribed[49-51].

Institute of Human Genetics, Erlangen: Exome sequencing was performed on a HiSeq 2500 (Illumina) platform with 125 bp paired-end sequencing using SureSelect v.5 capturing reagents (Agilent).

Radboud University Medical Center, Nijmegen: Exome sequencing was undertaken using Agilent SureSelectXT Human All Exon 50 Mb Kit, with sequencing on SOLiD 5500XL, producing an average sequence depth of 91X and average coverage of at least 20X for 89% of targets. For calling and annotation of variants, a custom in-house diagnostic pipeline was deployed[52].

UCL-Institute of Neurology, London: Exome sequencing was performed using Illumina's Nextera Rapid Capture. Indexed and pooled libraries were sequenced on Illumina's HiSeq3000 (100bp, paired-end). Reads were aligned with Novoalign. Duplicate read removal, format conversion, and indexing were performed with Picard. GATK was used to recalibrate base quality scores, perform local realignments around possible indels, and to call (HaplotypeCaller) and filter (VQSR) variants[46]. Annotated variant files were generated using ANNOVAR[53].

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

Sanger Sequencing for Variant Validation and Gene Screening

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

Enrichment Analysis

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

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

(3) CSF Neurotransmitter Analysis

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

(4) Comparative Modelling

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

(5) Histone Methylation Assays

H3K4 Methylation

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

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

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

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

Immunoblot Analysis of Histone Methylation

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

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

(6) RNA and Protein Measurements

Fibroblast RNA Extraction and cDNA Synthesis

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

Generation of Dopaminergic Neurons, RNA Extraction, cDNA Synthesis

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

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

PCR Analysis

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

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

Fibroblast Protein Preparation and Immunoblot Analysis

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

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

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

CSF Immunoblotting

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

(7) Statistics

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

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

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

References

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

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

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