European Journal of Human Genetics (2015) 23, 753–760

www.nature.com/ejhg

ARTICLE

Loss-of-function variants of *SETD5* cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome

Alma Kuechler[1,11], Alexander M Zink[2,11], Thomas Wieland[3], Hermann-Josef Lüdecke[1], Kirsten Cremer[2], Leonardo Salviati[4], Pamela Magini[5], Kimia Najafi[6], Christiane Zweier[7], Johanna Christina Czeschik[1], Stefan Aretz[2], Sabine Endele[7], Federica Tamburrino[8], Claudia Pinato[4], Maurizio Clementi[4], Jasmin Gundlach[3], Carina Maylahn[2], Laura Mazzanti[8], Eva Wohlleber[2], Thomas Schwarzmayr[3], Roxana Kariminejad[6], Avner Schlessinger[9], Dagmar Wieczorek[1], Tim M Strom[3,12], Gaia Novarino[10,12] and Hartmut Engels*[,2,12]

Intellectual disability (ID) has an estimated prevalence of 2–3%. Due to its extreme heterogeneity, the genetic basis of ID remains elusive in many cases. Recently, whole exome sequencing (WES) studies revealed that a large proportion of sporadic cases are caused by *de novo* gene variants. To identify further genes involved in ID, we performed WES in 250 patients with unexplained ID and their unaffected parents and included exomes of 51 previously sequenced child–parents trios in the analysis. Exome analysis revealed *de novo* intragenic variants in SET domain-containing 5 (*SETD5*) in two patients. One patient carried a nonsense variant, and the other an 81 bp deletion located across a splice-donor site. Chromosomal microarray diagnostics further identified four *de novo* non-recurrent microdeletions encompassing *SETD5*. CRISPR/Cas9 mutation modelling of the two intragenic variants demonstrated nonsense-mediated decay of the resulting transcripts, pointing to a loss-of-function (LoF) and haploinsufficiency as the common disease-causing mechanism of intragenic *SETD5* sequence variants and *SETD5*-containing microdeletions. *In silico* domain prediction of *SETD5,* a predicted SET domain-containing histone methyltransferase (HMT), substantiated the presence of a SET domain and identified a novel putative PHD domain, strengthening a functional link to well-known histone-modifying ID genes. All six patients presented with ID and certain facial dysmorphisms, suggesting that *SETD5* sequence variants contribute substantially to the microdeletion 3p25.3 phenotype. The present report of two *SETD5* LoF variants in 301 patients demonstrates a prevalence of 0.7% and thus *SETD5* variants as a relatively frequent cause of ID.
European Journal of Human Genetics (2015) **23,** 753–760; doi:10.1038/ejhg.2014.165; published online 20 August 2014

INTRODUCTION

Intellectual disability (ID, IQ<70) affects up to 3% of the general Western population.[1,2] Its underlying cause still remains elusive in a large proportion of cases due to its high locus heterogeneity. Recently, whole exome sequencing (WES) led to the identification of several novel genes mutated in ID.[3,4] Currently, the most successful approach in this context is the identification of *de novo* loss-of-function (LoF) sequence variants in a candidate gene in multiple patients. Due to the rarity of these variants, the occurrence of two or more independent events generates significant evidence for disease association.[5,6]

Distal deletions of the short arm of chromosome 3 were first characterised by cytogenetic and FISH analyses. Despite its rarity, the 3p- syndrome is well-described clinically. It is characterised by developmental delay/ID, microcephaly, low birth weight, growth retardation, hypotonia and craniofacial dysmorphisms.[7,8] Associated malformations such as congenital heart disease and polydactyly have been reported.[8,9] As in many other microdeletions, the contribution of the various genes in the deleted interval to the syndrome remained unclear until recently. For some microdeletion syndromes, haploinsufficiency of a single gene has been shown to be causal for the specific phenotype, such as *EHMT1* in terminal 9q deletion or *MEF2C* in microdeletion 5q14.3 syndrome.[10,11] Chromosomal microarray analysis (CMA) has allowed the precise characterisation of the deleted intervals in 3p- syndrome, and recently patients with small interstitial deletions of 3p25.3 ranging from 643 kb to 1.6 Mb in size have been described.[9,12–14] These patients showed many key features of the 3p- syndrome and their minimal deletion overlap comprised only three annotated genes, that is, *THUMPD3, SETD5* and *SETD5-AS1*. It has therefore suggested that these genes have a crucial role in the 3p25.3 deletion phenotype.[13] In addition, a recent

[1]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany; [2]Institute of Human Genetics, University of Bonn, Bonn, Germany; [3]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; [4]Department Salute Donna e Bambino, Clinical Genetics Unit, University of Padova, Padova, Italy; [5]U.O. Genetica Medica, Dipartimento d Scienze Mediche e Chirugiche (D.I.M.E.C.), Università di Bologna, Bologna, Italy; [6]Kariminejad Najmabadi Pathology and Genetics Center, Shahrak Gharb, Tehran, Iran; [7]Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; [8]Ambulatorio di Auxologia, Sindromologia e Malattie Rare, Dipartimento di Scienze Mediche e Chirurgiche (D.I.M.E.C.), Università di Bologna, Policlinico S.Orsola-Malpighi, Bologna, Italy; [9]Department of Pharmacology and Systems Therapeutics, and Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, USA; [10]Institute of Science and Technology Austria, Klosterneuburg, Austria
*Correspondence: Dr H Engels, Institute of Human Genetics, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany. Tel: +49 228 287 51055; Fax: +49 228 28751011; E-mail: hartmut.engels@uni-bonn.de
[11]These authors contributed equally to this work.
[12]These authors contributed equally as senior authors.
Received 11 April 2014; revised 3 July 2014; accepted 9 July 2014; published online 20 August 2014

copy number variations (CNV) study on autism spectrum disorder defined a minimal region of CNV overlap including SETD5 only.[15]

Here, we report two intragenic *SETD5* LoF sequence variants, one of them previously published,[3] as well as four non-recurrent microdeletions encompassing *SETD5,* substantially increasing the number of reported microdeletions. As also proposed in a recent report on further carriers of *SETD5* variants,[16] the phenotypic similarities between individuals carrying intragenic *SETD5* sequence variants and *SETD5* microdeletion carriers indicate that *SETD5* LoF variants alone cause an ID, for example, certain facial dysmorphisms, and that *SETD5* significantly contributes to the microdeletion 3p25.3 phenotype.

MATERIALS AND METHODS

Subjects

Written informed consent to the study and to the publication of the clinical photographs was obtained from the legal representatives of each participant. All investigations were performed in accordance with the Declaration of Helsinki and were approved by the respective local institutional review board. The retrospective reports of the microdeletions detected during routine CMA diagnostics did not require ethics committee approval.

We enrolled 250 individuals in our WES study. The inclusion criterion was developmental delay/ID with or without additional features (for example, craniofacial dysmorphism, organ malformation etc.), which could not be attributed to a clinically recognisable syndrome by an experienced clinical geneticist. Clinically relevant chromosomal aberrations were excluded by CMA analysis. Detailed clinical descriptions of the patients with *SETD5* variants are provided in Table 1 and in the clinical case reports (Supplementary Material). Exomes of 51 previously sequenced child–parents trios were included in the analysis.[3] All data interpretation was based on the GRCh37/hg19 human genome assembly.

Whole exome sequencing

Genomic DNA was extracted from peripheral blood samples of the affected individuals and their parents using DNA standard extraction kits. Exomes were enriched in solution and indexed with versions 3[3] and 5 of the SureSelect XT Human All Exon 50 Mb kit (Agilent Technologies, Santa Clara, CA, USA). Sequencing was performed as 101 bp paired-end reads on HiSeq2000/2500 systems (Illumina, San Diego, CA, USA). For patients 1 and 2, 9 and 14 Gb of sequence were generated, respectively, resulting in an average depth of coverage of 185 and 120, with 92 and 96% of the target regions covered at least 20 times. Image analysis and base calling was performed using Illumina Real Time Analysis. Reads were aligned against the human assembly hg19 (GRCh37) using BWA v 0.5.9 (http://bio-bwa.sourceforge.net/). Variant calling was performed specifically for the regions targeted by the exome enrichment kit using SAMtools (v 0.1.18; http://samtools.sourceforge.net/), PINDEL (v 0.2.4t; http://gmt.genome.wustl.edu/pindel/current/), ExomeDepth (v 1.0.0; http://www.stats.bris.ac.uk/R/web/packages/ExomeDepth/) and custom scripts. Variant quality was determined using the SAMtools varFilter script with default parameters except for the maximum read depth (−D) and the minimum *P*-value for base quality bias (−2), which were set to 9999 and 1e −400, respectively. In addition, a custom script was applied to mark all variants with adjacent bases of low median base quality. All variants were then annotated using custom Perl scripts. Annotation included known transcripts (UCSC Known Genes and RefSeq genes), known variants (dbSNP v 135; http://www.ncbi.nlm.nih.gov/SNP/), type of mutation and—if applicable—amino acid changes. The annotated variants were integrated into an in-house database. To discover putative *de novo* variants, we queried the database to show only those variants of a child that were not found in the parents. To reduce false positives, we filtered out variants that were already present in the database, or had a variant quality of <30, or did not pass the filter scripts. We then manually checked the raw read data of the remaining variants using the Integrative Genomics Viewer.

As a first step in the ongoing project, *de novo* variants in novel genes are being verified by Sanger sequencing. Depending on the nature of the variants, different prediction algorithms such as SIFT (http://sift.jcvi.org/), Polyphen2 (http://genetics.bwh.harvard.edu/pph2/) or Mutationtaster (www.mutationtaster.org) are applied to determine potential pathogenicity. Variants which are not analysable by prediction algorithms (such as the 81 bp deletion presented here) are examined individually. *SETD5* is the first gene identified in the cohort with two verified *de novo,* potentially deleterious variants.

Validation by sanger sequencing (Patients 1 and 2)

Bidirectional Sanger sequencing of *SETD5* was performed on whole blood genomic DNA in patients 1 and 2 and their respective parents using the ABI BigDye Terminator v.3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA, USA) to verify the variants and their *de novo* status. The 81 bp deletion was sequenced using a 3130xl Genetic Analyzer and the nonsense variant using an automated capillary sequencer ABI 3730 as reported previously (Life Technologies). Primer sequences are available upon request. Both variants as well as all other verified *de novo* SNVs were submitted to the Leiden Open Variation Database LOVD v.3.0 (http://databases.lovd.nl/shared/individuals/00016308 and /00016309). *SETD5* variants were annotated based on the transcript number NM_001080517.1.

CMA analyses

The CMA analyses of patients 3–6 were performed in different centres and on different platforms as part of routine diagnostics. Patients 3–6 were not part of the WES cohort described above. Details of the CMA analyses are summarised in Supplementary Table 1. Confirmation of CMA results and segregation analysis was carried out either by an independent array analysis (patients 3–5) or by qPCR (patient 6). qPCR for patient 6 was performed for the *ITGR1* locus, probes were designed with the Universal ProbeLibrary Assay Design Tools (Roche Applied Science, Penzberg, Germany).

In silico domain prediction

We examined structural features of SETD5 using various sequence analysis tools, including methods to predict secondary structure elements, and domain assignment approaches, including fold recognition methods and domain annotation databases (Supplementary Information). This approach, which was previously found useful in identifying structural similarities among related proteins,[17,18] increases the confidence of domain assignments.

CRISPR/Cas9 mutation modelling

Cas9 is a double-stranded DNA endonuclease that is guided to the cleavage site by a short (20 bp) guide RNA sequence. The guide RNA must be complementary to the target DNA, which must immediately precede a Cas9-specific protospacer-adjacent motif (PAM).[19] The resulting double-stranded break at the target site is repaired by the non-homologous end-joining (NHEJ) DNA repair pathway, which usually leads to the loss or introduction of additional bases. Guide sequences were designed using the online CRISPR Design tool (http://tools.genome-engineering.org). To co-express the single guide RNA (sgRNA) with Cas9, the sgRNA was cloned in the pSpCas9(BB) vector, as previously described.[20] One day prior to transfection, 1.3×10^5 HEK-293 cells were plated in 24-well plates. Cells were transfected using Lipofecatimine 2000 (Life Technologies) reagent according to manufacturer's instructions. To control transfection efficiency, one well was transfected with a pEGFP vector. After 48 h, 80–90% of the cells were GFP positive (data not shown). Two days after the transfection, cells were dissociated and used for isolation of clonal cell populations[20] or plated in 25 cm flasks for RNA preparation.

Semi-quantitative PCR and expression analysis

Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany). A total of 1 μg RNA was transcribed to cDNA using the SuperScriptIII cDNA kit (Life Technolgies) with random hexamers. Primers for SETD5 expression analysis were designed to amplify the region spanning from exon 5 to 7 of mouse SETD5 cDNA (transcript ENSMUST00000113155) (Figure 2). Primers for SETD5 expression analysis were designed to amplify the region spanning exons 6–8 of mouse SETD5 cDNA (transcript ENSMUST00000113157), corresponding to exons 6–8 in the human transcript depicted in Figure 2a (RefSeq transcript NG_034132.1). Pictures of the PCR products run on a 1.5% agarose gel were taken with a Quantum multi-imaging system (PEQLAB, Erlangen, Germany). Data were normalised to the density of GAPDH.

Table 1 Clinical data of patients 1–6 and of the patients described in previous reports

							Gunnarsson and Foyn			
	Patient 1	*Patient 2*[a]	*Patient 3*	*Patient 4*	*Patient 5*	*Patient 6*	*Bruun*[12]	*Peltekova et al*[9]	*Riess et al*[14]	*Kellogg et al*[13]
SETD5 aberration	Mutation chr3:g.9477570_9477650del	Mutation chr3:g.9490270C>T	Deletion 148 kb 4 genes	Deletion 371 kb 10 genes	Deletion 2.45 mb 46 genes	Deletion 11.16 Mb 71 genes	Deletion 1.6 mb 24 genes	Deletion 643 kb 12 genes	Deletion 1.24 Mb 11 genes	Deletion 684 kb 7 genes
Age (years) at last examination	20	9 $^{10}/_{12}$	7 $^{10}/_{12}$	1 $^{4}/_{12}$	6	2 $^{4}/_{12}$	4	22	3	11
Gender	F	F	M	M	F	F	F	F	F	F
Descent	Caucasian	Caucasian	Caucasian	Caucasian	Iranian	Caucasian	Caucasian	Caucasian	Caucasian	Chinese
Uneventful pregnancy	+	+	Imminent abortion, PROM	Caesarean section (altered umbilical artery flow)	+	+	After ICSI	Polyhydramnios, nausea, vomiting, cramping	+	+
Gestational weeks	40	40	34	31	40	39	37	37	39	40
Birth weight (g)/ (SD)	3210	3070	2190	1200	2550	2710	2440	2600	2990	3348
	−0.62	−0.95	−0.36	−1.27	−2.1	−1.51	−1.74	−0.86	−0.84	−0.30
Birth length (cm)/ (SD)	51	52	43	37	NR	50	45	45	49	53.3
	−0.32	0.14	−1.1	−1.45		−0.5	−1.87	−1.87	−0.95	0.72
Birth OFC (cm)/ (SD)	35	33	32.5	26.4	NR	33	32	33	35	NR
	−0.08	−1.46	0.17	−1.97		−1.23	−1.77	−0.64	0.31	
Height (cm)/(SD) at last evaluation	178	134	113.9	73.5	100	81	NR	135	93	143.9
	2.26	−1.18	−2.77	−3.29	−3.72	−2.27		−5.0	−0.80	−0.06
Weight (kg) at last evaluation	67	28.5	22.3	8.3	20	9.6	NR	42.2	NR	28.6
BMI (kg/m^2)	21.1	15.9	17.2	15.4	20	14.6	—	23.1	—	13.8
	(50^{th})	(25^{th}–50^{th})	(∼75^{th})	(10^{th}–25^{th})	(∼97^{th})	(10^{th}–25^{th})		(50^{th}–75^{th})		(∼3^{rd})
OFC (cm)/(SD) at last evaluation	58	50.5	53.2	45	48	44	NR	52	50.5	52
	2.46	−1.26	0.24	−2.2	−2.4	−3.38		−1.64	0.79	−0.23
ID	+	+	+ +	+	+ + +	+ +	+ + +	+ + +	+	+
Walking age (years)	2–3	1 $^{8}/_{12}$	2 $^{3}/_{12}$	Not yet	Not yet	Not yet	3	6	2	NR
First words (years)	4	4	3 $^{6}/_{12}$	Not yet	Not yet	Not yet	Not yet	NR	2	NR
Current speech	Fluent	Fluent	Some words	Not yet	No speech at 6 years	Syllables	No speech at 4 years	Initial 5–6 words lost, non-verbal as adult	NR	NR
Muscle hypotonia	−	−	+	+	+ +	+	+	NR	+	+
Seizures	1 febrile at 8 years	−	+	1 febrile	+	−	+	+	−	−
Brain anomalies (MRI)	−	NI	−	−	NI	−	Asymmetry of thalamus	−	NI	−
Behavioural anomalies	Dominant in known, anxious in unknown situations	Mild attention deficit disorder	ADHD	−	−	−	Poor eye contact and interaction, sleep disturbances	−	−	Obsessive-compulsive disorder, autistic spectrum
Vision/eye findings	Myopia, astigmatism	Strabismus	Strabismus right hemianopsia		Strabismus	Hyperopia, strabismus	Strabismus	Microphthalmia	Bilateral convergent strabismus	Strabismus (operated on at 2 years)
Low anterior hairline	+	+	−	−	−	−	NR	+	−	+
Arched eyebrows	−	−	−	+	+	−	NR	−	−	(+)
Ptosis	−	−	−	+	+	+	+	+	−	−
Palpebral fissures	Normal	Mildly downslanting	Normal	Blepharophimosis	Upslanting	Upslanting, telecanthus	Normal	Downslanting, blepharophimosis	Normal	Normal
Nose: broad bridge, bulbous tip	+	+	+	+	+	+	+	+	+	+
Anteverted nares	+	+	+	+	+	+	+	+	+	+
Philtrum	Long	Long	Long	Long	Long	Long, broad	Long	Smooth	Long, prominent	Long
Downturned corners of the mouth	(+)	+	+	+	+	+	+	+	(+)	−
Low set/ malformed ears	−	+	−	+	−	+	+	+	−	+
Postaxial hexadactyly	−	−	Bilateral, hands	Left hand	−	−	−	Bilateral, hands and feet	−	−
Other hand anomalies	Long and thin fingers	Prominent finger joints, broad distal phalanges	−	−	Tapering fingers	Tapering fingers, oedematous back of hands	−	−	−	Proximally placed thumbs
Other feet anomalies	−	Sandal gaps	−	−	−	−	Bilateral overlap of 2nd and 4th toes	2–3 syndactyly	−	NR
Congenital heart defect	−	−	−	−	−	−	ASD,[b]	ASD I	−	−
Others		Recurrent infections, constipation at younger age	Bifid uvula, oral frenula, hyperkeratosis, (kyphosis)	Cryptorchidism, complications of premature birth	−	Right helical pit, cutis marmorata	−	Cleft palate, bowel malrotation, scoliosis	−	Prominent lobes, right ear pit

Abbreviations: ADHD, attention deficit hyperactivity disorder; ASD, atrial septal defect; BMI, body mass index; F, female; ICSI, intracytoplasmatic sperm injection; ID, intellectual disability; M, male; MRI, magnetic resonance imaging; NI, not investigated; NR, not recorded; OFC, occipito-frontal circumference; PROM, premature rupture of membranes; SD, standard deviation; SETD5, SET domain-containing 5.

[a]Patient reported previously as ER14209 (3).

[b]Ventricular septal defect and a horizontal membrane dividing the atrium into two parts (suggestive of cor triatrium) and a wide coronary sinus.

Genomic coordinates are given based on the human assembly hg19 (GRCh37) and variant descriptions are based on transcript NM_001080517.1.

RESULTS

Clinical phenotype of *SETD5* sequence variant and microdeletion carriers

Clinical details of the six individuals presented here are provided in Table 1 and the clinical case reports in the Supplementary Data. All six patients displayed similar facial dysmorphisms in childhood such as mild hypertelorism, a tubular nose with prominent columella, anteverted nares and a broad nasal tip, a long philtrum and downturned corners of the mouth (Figure 1). In all but one, measurements at birth were normal. However, all four microdeletion carriers developed short stature and three of them also developed microcephaly during the first year of life. Neither of the patients with intragenic variants had microcephaly or short stature at birth or later in life and their ID was estimated to be mild to moderate. They showed a severely delayed speech development (first words at the age of 4 years), but eventually accomplished use of fluent sentences. ID and speech impairment was more pronounced in the patients with microdeletions. In addition, all four microdeletion carriers presented with muscular hypotonia and two of them had a history of seizures. Further, in two patients febrile seizures were noted. No internal malformations were reported. However, two microdeletion patients presented with postaxial hexadactyly of the hands.

SETD5 sequence variants

Analysis of the exome sequencing data of 301 child–parent trios identified intragenic *de novo SETD5* variants in two patients. No other genes with potentially pathogenic variants were present in patients 1 and 2. The *SETD5* variants were identified in a cohort of 250 patients with ID of unknown origin (patient 1) and a previously published cohort of 51 individuals with ID (patient 2, published with limited clinical description).[3] Both variants were confirmed by Sanger sequencing. Patient 1 carried a *de novo* 81 bp deletion which deleted 7 codons of exon 7 and 60 adjacent base pairs of intron 7 (chr3:g.9477570_9477650del) (Figure 1). The four nucleotides upstream of the 5′ deletion breakpoint and the last four nucleotides of the deleted segment were identical (AGCA) and thus constituted a junctional microhomology[21] (Figure 1k, blue bar). In addition to the deletion, a *de novo* nucleotide substitution (chr3:g.9477546A>G, c.523A>G, p.(Ser175Gly)) was detected 25 nucleotides upstream of the 5′ deletion breakpoint (Figure 1k). This substitution was only present on WES reads that also carried the deletion (Supplementary Figure 1), which was verified by PCR (data not shown). Thus, any independent effect of the substitution is most probably abolished by a nonsense-mediated mRNA decay (NMD) of the allele (see below). Patient 2 had previously been found to carry a *de novo* nonsense variant[3] (chr3:g.9490270C>T, c.2302C>T, p.(Arg768*); Figure 1l), which was predicted to cause NMD (www.mutationtaster.org).[22]

Further targeted copy number analysis of exome sequence data in the 301 exomes revealed no deletion encompassing *SETD5*. However, diagnostic molecular karyotyping performed at four different centres detected one terminal and three interstitial microdeletions affecting 3p25.3. The sizes of the interstitial deletions in 3p25.3 ranged from 148 kb in individual 3 (containing four genes) and 371 kb (individual 4, 10 genes) to 2.45 Mb (individual 5, 46 genes). The terminal deletion in individual 6 comprised 11.16 Mb and 71 genes. The four deletions had a common minimal region in 3p25.3 (chr3:9,422,487_9,542,885). This region encompassed the entire *SETD5* gene and parts of *Homo sapiens* THUMP domain-containing 3 (*THUMPD3*) and *LHFPL4* (Figure 1m) as well as the non-coding antisense RNA *SETD5-AS1*.

In silico domain prediction

SET domain-containing proteins have been shown to possess histone H3-specific methyltransferase activity.[23] SETD5 is a 1442 amino acid protein, highly expressed in the cerebral cortex, the intestine and the eye (Figure 2a), that is predicted to contain a SET domain. To date, however, no methyltransferase activity has been demonstrated. Our *In silico* domain prediction analyses (details: Supplementary Information) suggested that SETD5 is a multidomain protein containing a SET domain and a putative PHD domain which has not been described before for SETD5 (Supplementary Figure 2). Both PHD and SET domains are often found in nuclear proteins that interact with chromatin,[24,25] further supporting a possible role of SETD5 in chromatin modification.

CRISPR/Cas9 mutation modelling

To substantiate haploinsufficiency as the disease-causing mechanism especially of the 81 bp deletion, we analysed the effect of both intragenic variants on the *SETD5* mRNA stability *in vitro*. Since a second sample for RNA analysis was declined for patient 1, we decided to mimic the mutations employing the CRISPR/Cas9 system.[26] To guide the double-stranded DNA endonuclease Cas9 to the cleavage site, short (20 bp) guide RNA sequences are necessary. To study the effect of the 81 bp splice site deletion (Figure 2b), HEK-293 cells were co-transfected with two independent sequences, denoted 'guide 1' and 'guide 2', which targeted Cas9-mediated cleavage to the proximity of the 5′ and the 3′deletion breakpoints, respectively (Figure 2c left panel and Figure 2d). To mimic the effect of the nonsense variant on mRNA stability, we used a guide sequence that directed the cleavage to the precise location of the identified C to T substitution. Analyses of the *SETD5* transcripts by semi-quantitative PCR 1 week after transfection revealed a significantly lower amount of *SETD5* transcript both in cells carrying the Cas9-mediated deletion (Figure 2d) and carrying the nonsense variant (Figure 2c, right panel) compared with the controls. Thus, our analyses provided support for the hypothesis that NMD is the consequence of both intragenic variants and that haploinsufficiency is the disease-causing mechanism common to the microdeletions and intragenic sequence variants.

DISCUSSION

We report on six individuals with different types of SETD5 mutations, that is, two intragenic sequence variants detected by WES and four overlapping microdeletions affecting SETD5 detected by diagnostic microarray analyses.

Individuals with *SETD5* intragenic sequence variants and microdeletions present with a similar craniofacial phenotype

All six patients described in the present report, the seven patients with LoF variants published recently[16] and the four patients with overlapping microdeletions described in previous studies[9,12–14] (Table 1) presented with a characteristic craniofacial phenotype. This comprised a tubular nose with broad nasal bridge and bulbous nasal tip, anteverted nares, a long philtrum and downturned corners of the mouth (Figures 1a–j) or a thin upper lip.[16] Striking eyebrows with variable morphology (full, straight, arched, broad and synophrys) are often present. Other clinical signs show incomplete penetrance and are present in comparable frequencies in LoF sequence variant and microdeletion carriers: Congenital heart defects were present in two of the eight microdeletion carriers[9,12] and in two of the nine carriers of intragenic sequence variants. Postaxial hexadactyly was observed in three microdeletion patients (ie patients 3 and 4 from the present report and one previously published patient[9]) and in one of the LoF

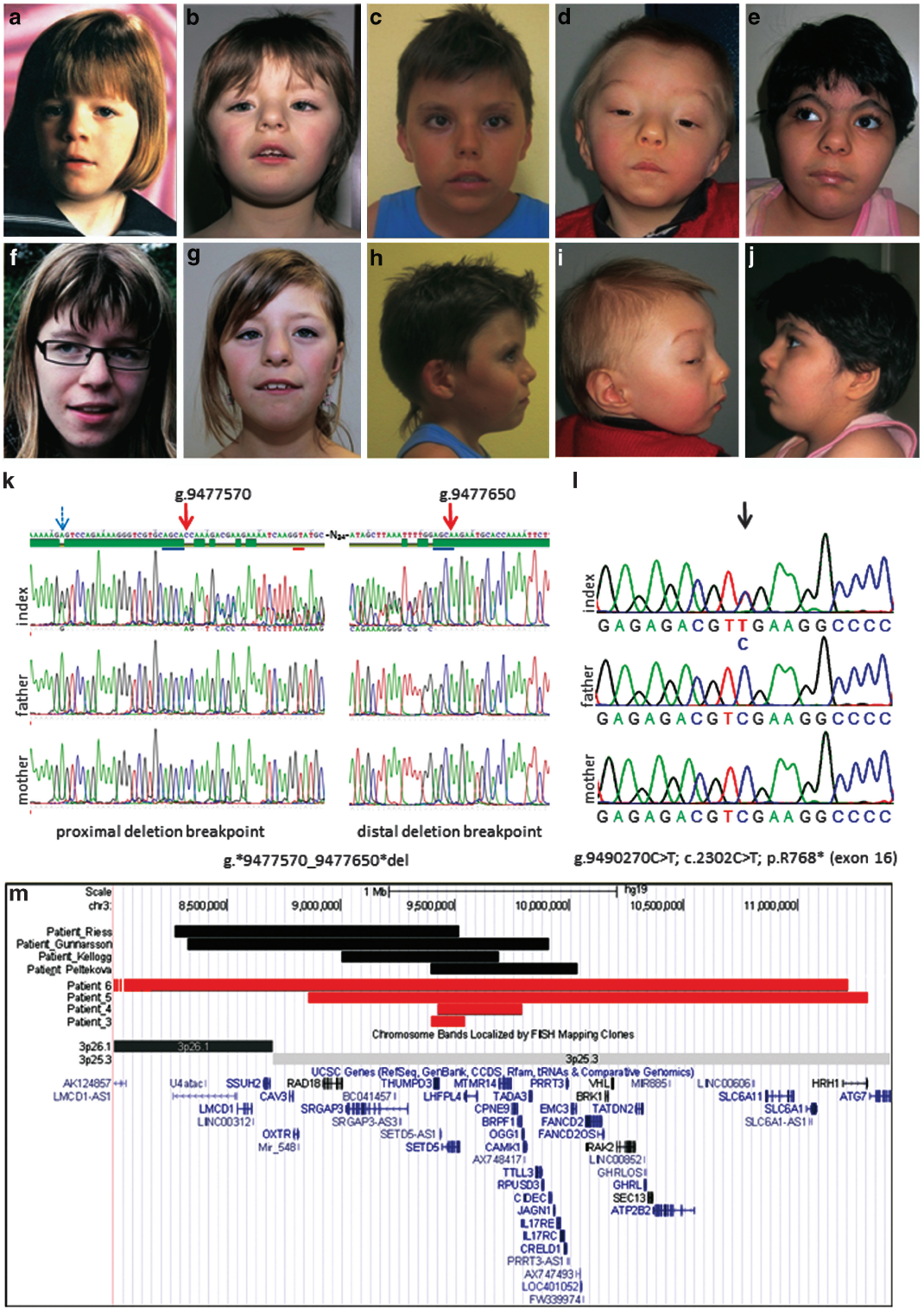

Figure 1 Facial phenotypes of individuals with a *SETD5* sequence variant or microdeletion, results of the mutation analysis and the size and localisation of overlapping deletions. (**a**–**j**) Facial phenotypes of patients 1–5 (consent for the publication of photographs of patient 6 was withdrawn). All six patients share certain facial phenotypic features, in particular a broad nasal bridge, anteverted nares, a long philtrum and downturned corners of the mouth. (**a**, **f**) show patient 1 who carries an 81 bp deletion (Figure 1k) at different ages; (**b**, **g**) show patient 2 who carries a nonsense variant (Figure 1l)) at different ages; (**c**, **h**) patient 3; (**d**, **i**) patient 4 and (**e**, **j**) patient 5. (**k**, **l**) electropherograms of the variant verifications by Sanger sequencing in patients. (**k**) Scheme of the 81 bp *de novo* deletion identified in patient 1. The first and last nucleotides of the deletion are marked with red arrows, and the last four nucleotides before the deletion (AGCA) and the last four nucleotides of the deleted segment (also AGCA) are underlined in blue. The splice-donor site (GT) of intron 7 is underlined by a red bar. An additional *de novo* base substitution (A>G leading to AGT>GGT, Ser>Gly) is indicated by a blue dashed arrow (details: Supplementary Figure 1). (**l**) *de novo* nonsense base substitution found in patient 2. (**m**) Scheme of overlapping deletions of patients 3–6 compared with those previously published in the literature.

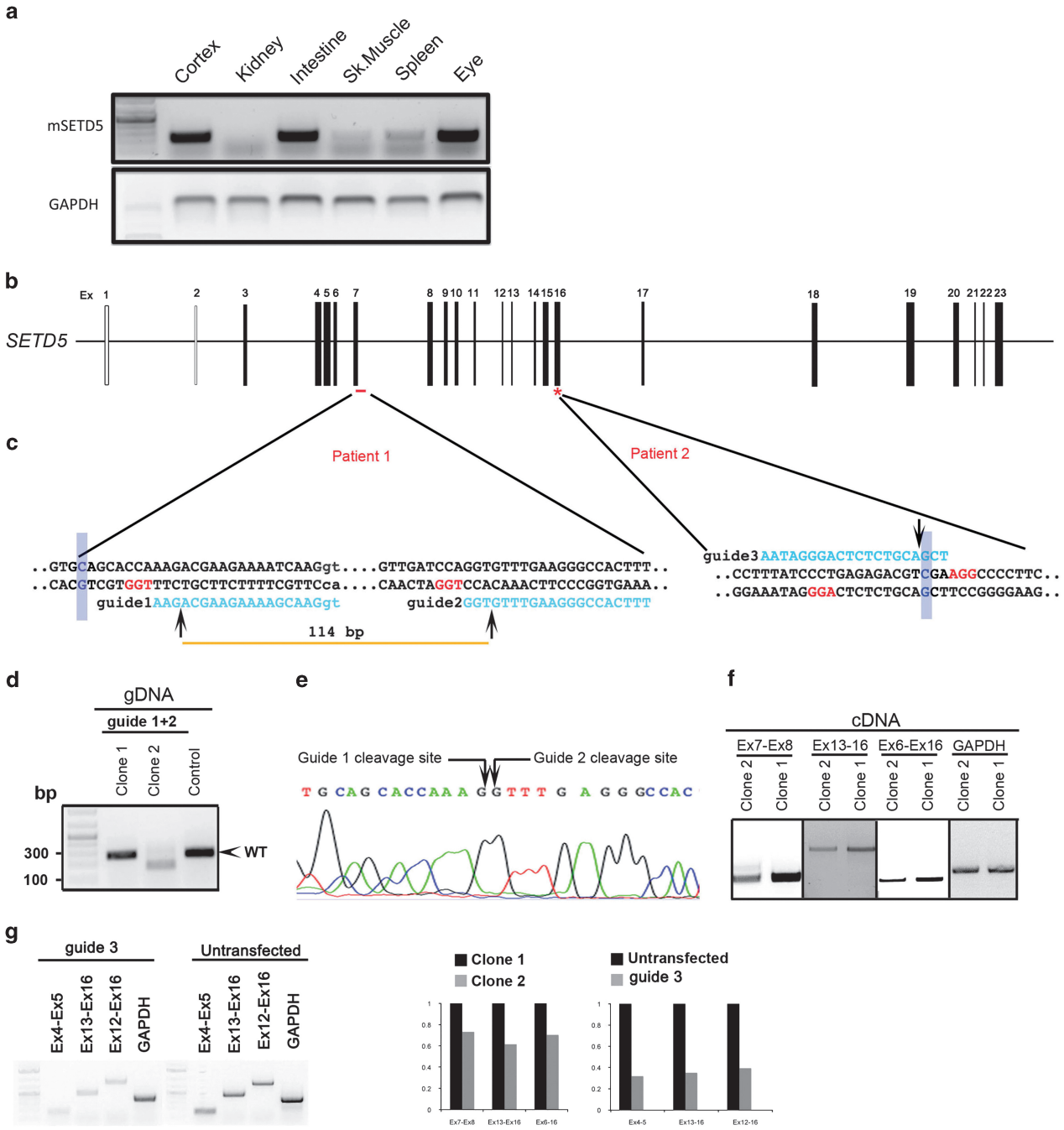

Figure 2 *SETD5* expression and Cas9-mediated genome editing. (**a**) SETD5 expression was probed in adult mouse tissues by semi-quantitative PCR. SETD5 expression is high in the cerebral cortex, the intestine and the eye. Significantly lower expression was observed in the other tissues including the kidney (at higher acquisition exposure), the skeletal muscle and the spleen. (**b–g**) Cas9-mediated genome editing simulating the variants of Patients 1 and 2: (**b**) *SETD5* gene structure. Exon (Ex) numbering corresponds to RefSeq transcript NG_034132.1. (**c**) Schematic representation of the Cas9-endonuclease targeting sites. The PAM sequence is labelled in red. Black arrows indicate RNA-guided cleavage sites. To mimic the deletion of Patient 1, Cas9-endonuclease was directed by guide 1 and guide 2 (left panel). The guide 1-mediated cutting site is located 6 bp downstream of the 5′ deletion breakpoint (purple box), while the guide 2 cutting site is located 44 bp downstream of the 3′ breakpoint. Cells transfected with both guide 1 and guide 2 are predicted to carry a deletion of ~114 bp. For the nonsense variant in patient 2, guide 3 which surrounds the mutated base was used (right panel, purple box). (**d**) PCR using primers flanking the predicted deletion region of genomic DNA (gDNA) showing the successful deletion in clone 2 of HEK-293 cells transfected with guide 1 + guide 2. The wt band was predicted to be 274 bp, clone 2 shows a band of ~150 bp. (**e**) gDNA sequence of clone 2 showing the 114 bp deletion. (**f–g**) messenger RNA stability was probed by semi-quantitative PCR using three separate primer pairs located in different exons. GAPDH was used as internal control. Untransfected cells and clone 1 were used as controls. Band intensity quantification is shown on the right. Quantification and visual inspection of band intensity show a significant reduction of SETD5 transcript level both for guide 1/guide 2 and guide 3 in comparison to controls.

sequence variant carriers.[16] Almost all LoF variant carriers and three of the microdeletion carriers show some degree of behavioural abnormalities (for example, attention deficit, autistic features, obsessive-compulsive disorder and ritualised behaviour). Together with ID, these findings seem to constitute a core phenotype for *SETD5* mutations. Patients with larger deletions displayed additional facial findings—five had ptosis, and four had either blepharophimosis or an abnormal slant of palpebral fissures. The clinical similarities between *SETD5* LoF sequence variant and microdeletion carriers suggest that *SETD5* contributes substantially to the core phenotype of the 3p25.3 microdeletion syndrome in addition to causing ID.

The growth pattern of the carriers of intragenic *SETD5* variants seems to be different from the growth pattern of the microdeletion patients. All individuals were born with normal head circumference, weight and length (except for the low birth weight of our patient 5 and of patient 7 of Grozeva *et al*[16]). While all four of the present and one of the published[9] microdeletion carriers developed short stature (total: five out of eight), the same is true only for one out of five LoF variant carriers with information on postnatal height (no height information available for four of the seven recently published patients[16]). While three of the four microdeletion carriers presented here developed microcephaly postnatally, all nine carriers of *SETD5* LoF variants remained normocephalic. This suggests that *SETD5* variants *per se* may have no influence on the development of microcephaly.

Several other clinical differences were observed between the carriers of intragenic sequence variants and microdeletions: while developmental delay was present in all individuals, only the microdeletion patients presented with muscular hypotonia. Furthermore, the nine individuals with LoF variants spoke their first words between 1 and 4 years of age and can now talk and communicate their needs (our patients 1 and 2 speak fluently in complete sentences now). In contrast, all microdeletion patients characterised for this clinical sign showed more severe speech impairments.

The microdeletions of the four patients presented here further narrow down the smallest region of microdeletion overlap to 94 kb (chr3:9,422,487_9,516,586). It affects only two protein-coding RefSeq genes: *SETD5* and parts of *THUMPD3*. This may point to an aetiological role of the second gene *THUMPD3* (*Homo sapiens* THUMP domain-containing 3), a predicted RNA-methyltransferase. Haploinsufficiency of this gene may lead to the clinical signs that appear exclusive to patients with microdeletions. However, effects, for example, of the long non-coding RNA *SETD5-AS1* which is also localised in the minimal region cannot be excluded.

Mutational spectrum of *SETD5*

CRISPR/Cas9 modelling of both the 81 bp deletion and the nonsense variant (patients 1 and 2) provided evidence that these *SETD5* variants trigger NMD and thus *SETD5* LoF. For the microdeletions including *SETD5*, haploinsufficiency may be assumed. Thus, loss of *SETD5* gene function and haploinsufficiency are likely to be the common aetiological mechanisms in the intragenic variants presented here and all microdeletions affecting *SETD5*. The 81 bp *de novo* deletion detected in patient 1 is interesting in two additional respects. First, this indel is too large to be detected by most variant calling algorithms that are routinely applied in WES studies such as SAMtools or GATK. This stresses the importance of algorithms aimed at the detection of larger indels such as the algorithm Pindel used in the present study. Second, the deletion is flanked by junctional microhomologies of four nucleotides. Several microhomology-mediated mechanisms have been proposed to explain the formation of CNVs since short stretches of microhomology have been detected at the majority of breakpoint sites of non-recurrent CNVs. This signature is consistent with mechanisms such as NHEJ, alternative NHEJ (also known as microhomology-mediated end joining), fork stalling and template switching, and microhomology-mediated break-induced replication.[21]

Putative truncating variants (frameshift or nonsense) of *SETD5* are in fact rare. None were identified in either the 1000 Genomes Project or in 2500 in-house control exomes. The Residual Variation Intolerance (RVI) score, which quantifies gene intolerance to functional mutations,[27] of SETD5 is −0.79 (12.59th percentile) and thus even lower than the average RVI score for developmental disorders (−0.56; 19.54th percentile). This suggests that the degree of intolerance to deleterious variants of *SETD5* is significantly more pronounced than in the average of genes known to cause developmental disorders. The 81 bp deletion in patient 1 was not present in dbvar (March 2014) and the nonsense variant in patient 2 was not present in dbSNP139.

SETD5—another epigenetic regulator causing ID

The phenotype caused by intragenic *SETD5* variants can be categorised within the group of 'neurodevelopmental disorders'[28] for which more than 500 causative genes have been identified already.[29] Recently, the epigenetic regulation of gene transcription in particular has emerged as a process of major importance in cognition and ID. Post-translational histone modification including methylation, acetylation, ubiquitination, phosphorylation and sumoylation[30] is a key epigenetic mechanism in transcription regulation. Kleefstra *et al*[31] categorised the known epigenetic genes underlying neurodevelopmental disorders into writers, erasers, chromatin remodelers of the DEAD/H ATPase helicase family and other readers/chromatin remodelers.

SETD5 probably is a new member of the 'writers' group of epigenetic ID genes. It encodes SET Domain-Containing Protein 5 and belongs to the SET methyltransferase family, which catalyse methylation of histone H3 and H4 lysine residues.[32] Extensive data on the HMT activities of SET domain proteins and their effects on gene regulation are available,[32] but detailed data on human SETD5 are lacking. Using a variety of computational sequence analysis tools, we substantiated the presence of the SET domain and newly described a putative PHD domain for SETD5. PHD domains are involved in the interaction of nuclear proteins with chromatin,[24,25] further suggesting a role of *SETD5* in chromatin modification. However, functional follow-up studies of SETD5 are warranted to elucidate SETD5 HMT activity and the molecular mechanisms underlying this disorder.

The potential HMT *SETD5* is thus added to the list of 'writer' ID genes, which includes several other HMT genes causing recognisable syndromic conditions involving ID such as *EHMT1* (Kleefstra syndrome), *MLL2* (Kabuki syndrome) and *NSD1* (Sotos syndrome). Based on a sample of patients with moderate to severe ID, rare *de novo* LoF variants in *SETD5* have recently been proposed to be a relatively frequent (0.7%) cause of ID.[16] The present finding of two LoF variants in a total of 301 patients (0.67%) representing the whole spectrum of ID patients (mild to severe ID, both with and without additional dysmorphisms and/or malformations) supports this conclusion and extends it to idiopathic ID in general.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to the patients and their families for participating in this study. We thank the 'German Mental Retardation Network' (MRNET) and Sabine Kaya, Daniela Falkenstein and Mike Liu for their excellent technical assistance. This work was supported in part by the German Ministry of Research and Education (grant numbers 01GS08164, 01GS08167, 01GS08163, German Mental Retardation Network) as part of the National Genome Research Network and by the European Commission's FP7 CHERISH programme (grant number 223692).

1 Flore LA, Milunsky JM: Updates in the genetic evaluation of the child with global developmental delay or intellectual disability. *Semin Pediatr Neurol* 2012; **19**: 173–180.
2 Ropers HH: Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* 2010; **11**: 161–187.
3 Rauch A, Wieczorek D, Graf E *et al*: Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012; **380**: 1674–1682.
4 de Ligt J, Willemsen MH, van Bon BW *et al*: Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012; **367**: 1921–1929.
5 He X, Sanders SJ, Liu L *et al*: Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet* 2013; **9**: e1003671.
6 Sanders SJ, Murtha MT, Gupta AR *et al*: *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
7 Malmgren H, Sahlen S, Wide K, Lundvall M, Blennow E: Distal 3p deletion syndrome: detailed molecular cytogenetic and clinical characterization of three small distal deletions and review. *Am J Med Genet A* 2007; **143A**: 2143–2149.
8 Shuib S, McMullan D, Rattenberry E *et al*: Microarray based analysis of 3p25-p26 deletions (3p- syndrome). *Am J Med Genet A* 2009; **149A**: 2099–2105.
9 Peltekova IT, Macdonald A, Armour CM: Microdeletion on 3p25 in a patient with features of 3p deletion syndrome. *Am J Med Genet A* 2012; **158A**: 2583–2586.
10 Kleefstra T, Brunner HG, Amiel J *et al*: Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am J Hum Genet* 2006; **79**: 370–377.
11 Zweier M, Gregor A, Zweier C *et al*: Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Hum Mutat* 2010; **31**: 722–733.
12 Gunnarsson C, Foyn Bruun C: Molecular characterization and clinical features of a patient with an interstitial deletion of 3p25.3-p26.1. *Am J Med Genet A* 2010; **152A**: 3110–3114.
13 Kellogg G, Sum J, Wallerstein R: Deletion of 3p25.3 in a patient with intellectual disability and dysmorphic features with further definition of a critical region. *Am J Med Genet A* 2013; **161A**: 1405–1408.
14 Riess A, Grasshoff U, Schaferhoff K *et al*: Interstitial 3p25.3-p26.1 deletion in a patient with intellectual disability. *Am J Med Genet A* 2012; **158A**: 2587–2590.
15 Pinto D, Delaby E, Merico D *et al*: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 2014; **94**: 677–694.
16 Grozeva D, Carss K, Spasic-Boskovic O *et al*: *De novo* loss-of-function mutations in setd5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet* 2014; **94**: 618–624.
17 Dokudovskaya S, Waharte F, Schlessinger A *et al*: A conserved coatomer-related complex containing Sec13 and Seh1 dynamically associates with the vacuole in Saccharomyces cerevisiae. *Mol Cell Proteomics* 2011; **10**: 006478.
18 van Dam TJ, Townsend MJ, Turk M *et al*: Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proc Natl Acad Sci USA* 2013; **110**: 6943–6948.
19 Cong L, Ran FA, Cox D *et al*: Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339**: 819–823.
20 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F: Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013; **8**: 2281–2308.
21 Ottaviani D, Lecain M, Sheer D: The role of microhomology in genomic structural variation. *Trends Genet* 2014; **30**: 85–94.
22 Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010; **7**: 575–576.
23 Rea S, Eisenhaber F, O'Carroll D *et al*: Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000; **406**: 593–599.
24 Bienz M: The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* 2006; **31**: 35–40.
25 Jenuwein T, Laible G, Dorn R, Reuter G: SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* 1998; **54**: 80–93.
26 Sander JD, Joung JK: CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014; **32**: 347–355.
27 Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB: Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet* 2013; **9**: e1003709.
28 American Psychiatric Association AAP. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Arlington, VA, USA: American Psychiatric Publishing, 2013.
29 van Bokhoven H: Genetic and epigenetic networks in intellectual disabilities. *Annu Rev Genet* 2011; **45**: 81–104.
30 Tan M, Luo H, Lee S *et al*: Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 2011; **146**: 1016–1028.
31 Kleefstra T, Schenck A, Kramer JM, van Bokhoven H: The genetics of cognitive epigenetics. *Neuropharmacology* 2014; **80**: 83–89.
32 Martin C, Zhang Y: The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 2005; **6**: 838–849.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)