Human Molecular Genetics, 2014, Vol. 23, No. 22 **5940–5949**
doi:10.1093/hmg/ddu318
Advance Access published on June 20, 2014

A missense mutation in the PISA domain of HsSAS-6 causes autosomal recessive primary microcephaly in a large consanguineous Pakistani family

Muzammil A. Khan[1,†], Verena M. Rupp[2,†], Meritxell Orpinell[4,†], Muhammad S. Hussain[5,6], Janine Altmüller[5,7], Michel O. Steinmetz[10], Christian Enzinger[3], Holger Thiele[5], Wolfgang Höhne[5], Gudrun Nürnberg[5], Shahid M. Baig[11], Muhammad Ansar[12], Peter Nürnberg[5,8,9], John B. Vincent[13], Michael R. Speicher[2], Pierre Gönczy[4] and Christian Windpassinger[2,*]

[1]Gomal Centre of Biochemistry and Biotechnology, Gomal University D.I.Khan, Khyber Pakhtoonkhuwa, Pakistan, [2]Institute of Human Genetics and [3]Department of Neurology, Medical University of Graz, Graz, Austria, [4]Swiss Institute for Experimental Cancer Research (ISREC), Swiss Federal Institute of Technology (EPFL), School of Life Sciences, Lausanne, Switzerland, [5]Cologne Center for Genomics (CCG), [6]Institute of Biochemistry I, Medical Faculty, [7]Institute of Human Genetics, [8]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD) and, [9]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany, [10]Laboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institut, Villigen PSI, Switzerland, [11]Human Molecular Genetics Laboratory, Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan, [12]Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan and [13]Molecular Neuropsychiatry and Development (MiND) Lab, The Campbell Family Brain Research Institute, The Centre for Addiction & Mental Health (CAMH), Toronto, Ontario, Canada

Received April 30, 2014; Revised and Accepted June 17, 2014

Asymmetric cell division is essential for normal human brain development. Mutations in several genes encoding centrosomal proteins that participate in accurate cell division have been reported to cause autosomal recessive primary microcephaly (MCPH). By homozygosity mapping including three affected individuals from a consanguineous MCPH family from Pakistan, we delineated a critical region of 18.53 Mb on Chromosome 1p21.3-1p13.1. This region contains the gene encoding HsSAS-6, a centrosomal protein primordial for seeding the formation of new centrioles during the cell cycle. Both next-generation and Sanger sequencing revealed a homozygous c.185T>C missense mutation in the *HsSAS-6* gene, resulting in a p.Ile62Thr substitution within a highly conserved region of the PISA domain of HsSAS-6. This variant is neither present in any single-nucleotide polymorphism or exome sequencing databases nor in a Pakistani control cohort. Experiments in tissue culture cells revealed that the Ile62Thr mutant of HsSAS-6 is substantially less efficient than the wild-type protein in sustaining centriole formation. Together, our findings demonstrate a dramatic impact of the mutation p.Ile62Thr on HsSAS-6 function and add this component to the list of genes mutated in primary microcephaly.

INTRODUCTION

Primary autosomal recessive microcephaly (MCPH; MIM 251200) is a rare heterogeneous developmental congenital brain disorder characterized by a reduced size of the cerebral cortex, which results in a smaller occipitofrontal circumference of the head that lies at least 3 standard deviations (SD) below the age, sex and ethnically matched mean (1–3). Primary

*To whom correspondence should be addressed at: Institute of Human Genetics, Medical University of Graz, Harrachgasse 21/8, 8010 Graz, Austria. Tel: +43 3163804114; Email: christian.windpassinger@medunigraz.at
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint first Authors.

microcephaly is present at birth but can already be diagnosed by 30 weeks of gestation and is nearly always accompanied by non-progressive intellectual disability (4–6). In some MCPH cases, reduced height and epilepsy have been reported, but besides the observation of simplified gyri, primary microcephaly is usually not accompanied by any other severe brain defects (4,5). Although rare in the non-consanguineous population (~1 in 1 Million), microcephaly has a higher incidence (1 in 100 000) in countries where endogamy is still an integral part of the tradition, such as Pakistan or Middle Eastern countries (6–9). In the past 10 years, twelve primary MCPH loci and genes have been reported, namely *MCPH1*, *WDR62*, *CDK5RAP2*, *CPAP/CENPJ*, *CASC5*, *ASPM*, *STIL*, *CEP63*, *CEP135*, *CEP152*, *PHC1*, *CDK6* and *ZNF335* (10–22). The vast majority of these genes encode centrosomal proteins, including five (CPAP, STIL, CEP63, CEP135, CEP152) that are required for efficient centriole formation. Mutations in these genes have an adverse effect on neuronal development, possibly by preventing proper asymmetric division of neuronal progenitor cells in the ventricular zone of the developing neocortex (23–25).

In progeny from consanguineous marriages suffering from an autosomal recessive trait, it is assumed that causative homozygous mutations are located in genomic regions that are homozygous-by-descent (HBD)—so called autozygous regions (26). HBD mapping using SNP arrays is an efficient genome-wide approach for identifying such autozygous regions (27–29). Subsequent massive parallel sequencing has made it cheaper and easier to screen these HBD regions for mutations (28).

In this study, we screened a large consanguineous Pakistani family with four patients diagnosed with autosomal recessive MCPH. After excluding all loci containing known microcephaly genes, we eventually identified in the affected family members a homozygous mutation in *HsSAS-6* (spindle assembly 6 homolog of *Caenorhabditis elegans*), a gene encoding the centrosomal protein HsSAS-6, which is critical for centriole formation and thus for proper cell division (30,31).

RESULTS

The family in this study was recruited from the urban area of Dera Ismail Khan within the Khyber Pakhtunkhwa province of Pakistan. The pedigree includes five generations with two affected girls of age 6 and 3.5 years (V-1 and V-3) (Fig. 1A) and two men of age 50 and 42 years (IV-7 and IV-8) (Fig. 1A), who are all progeny from consanguineous marriages (Fig. 2). All four affected individuals had a head circumference of up to 19.5 SD below the mean, which was combined with severe mental retardation and an IQ ranging between 20 and 40 (32). Accordingly, for all four patients, limited speech and pronunciation problems were reported. Otolaryngologic and ophthalmic examinations were normal in all, except for Patient V-1, who had a strabismus affecting the right eye. Strabismus has not been previously associated with microcephaly but is not uncommon in the otherwise healthy population (29). In contrast to the other affected family members, Patient V-1 was unable to walk by the age of 6 despite normal bone development. Epileptic seizures were reported for the two adult patients, Brothers IV-7 and IV-8, but not for the younger ones (see Table 1).

Computed tomography was performed on Patient V-3. This demonstrated microcephaly with poorly confined basal ganglia and missing delineation of the internal capsule. Moreover, abnormal formation of the lateral ventricles and a dysmorphic infratentorial region with hypoplasia of the vermis cerebella were also noted (Fig. 1B).

DNAs of Patients IV-7, V-1 and V-3 were sampled for genomic analysis, whereas that of Patient IV-8 could not be collected owing to aggressive behavior during sampling. Initially, we excluded all known MCPH loci in this family using short tandem repeat (STR) markers to demonstrate heterozygosity in the relevant genomic regions. We then performed a genome-wide linkage analysis assuming individuals II:5 and II:6 to be first cousins (see Fig. 2B). Genotype data were generated for three affected individuals (IV-7, V-1 and V-3) using the NspI 250K SNP array from Affymetrix. We observed a single significant peak with a maximum multipoint LOD score of 3.9 on Chromosome 1 (Supplementary Material, Fig. S1). The underlying homozygous region, shared between all affected individuals included in the analysis, comprised ~20 cM on the short arm of Chromosome 1 at cytoband 1p21.3-1p13.1 (Fig. 2A). The critical interval of the new MCPH locus is defined by the two SNP markers rs555557 at position 98,912,075 bp and rs2251406 at position 117,445,365 bp, spanning a region of 18.53 Mb according to UCSC human genome assembly hg19. The segregation of the interval was confirmed by genotyping all available family members with six highly polymorphic STR microsatellite markers located between 96.8 and 111.2 Mb on Chromosome 1p (Fig. 2B). A maximum two-point LOD score of 2.8 for marker D1S495 (102.56 Mb; hg19) and multipoint LOD score of 3.53 were calculated for the area between marker D1S2671 (101.27 Mb) and D1S495 (102.56 Mb) (Supplementary Material, Table S1).

As several known MCPH genes contribute to proper cell division, we screened by Sanger sequencing four candidate genes that were known to potentially contribute to this process and that are located within the homozygous region on Chromosome 1: *WDR47*, *PSRC1*, *NGF* and *HsSAS-6* (9,14,18,19,33–36). Only the sequence of *HsSAS-6* revealed a homozygous NM_194292.1: c.185T>C (Chr1:100588787A>G) transition in a highly conserved region in exon 3 in all patients under investigation (Fig. 3A and B). As a consequence, the hydrophobic non-polar isoleucine at position 62 in the HsSAS-6 protein is exchanged by a hydrophilic polar threonine (p.Ile62Thr). Consistent with the linkage analysis, all healthy family members that were carriers of the disease allele were heterozygous for the mutation, whereas individual V-2 was homozygous for the *HsSAS-6* wild-type allele (Fig. 3A). The mutation was not found in 116 unrelated, unaffected Pakistani individuals by Sanger sequencing. We further sequenced *HsSAS-6* in patients of 19 additional MCPH families from Pakistan, which were previously excluded for mutations in all known MCPH genes, but did not identify further mutations in *HsSAS-6*.

In order to exclude with certainty any other mutations in known microcephaly genes and in genes located in the HBD region in this family, we conducted whole-exome sequencing. We found another homozygous missense mutation, c.656C>T (p.Ser219Leu), close to the splice site of exon 8 of *CAPZA1* (F-actin-capping protein subunit alpha-1). This gene encodes a protein required for the regulation of actin polymerization (38). Neither the mutation in *HsSAS-6* nor the one in *CAPZA1*

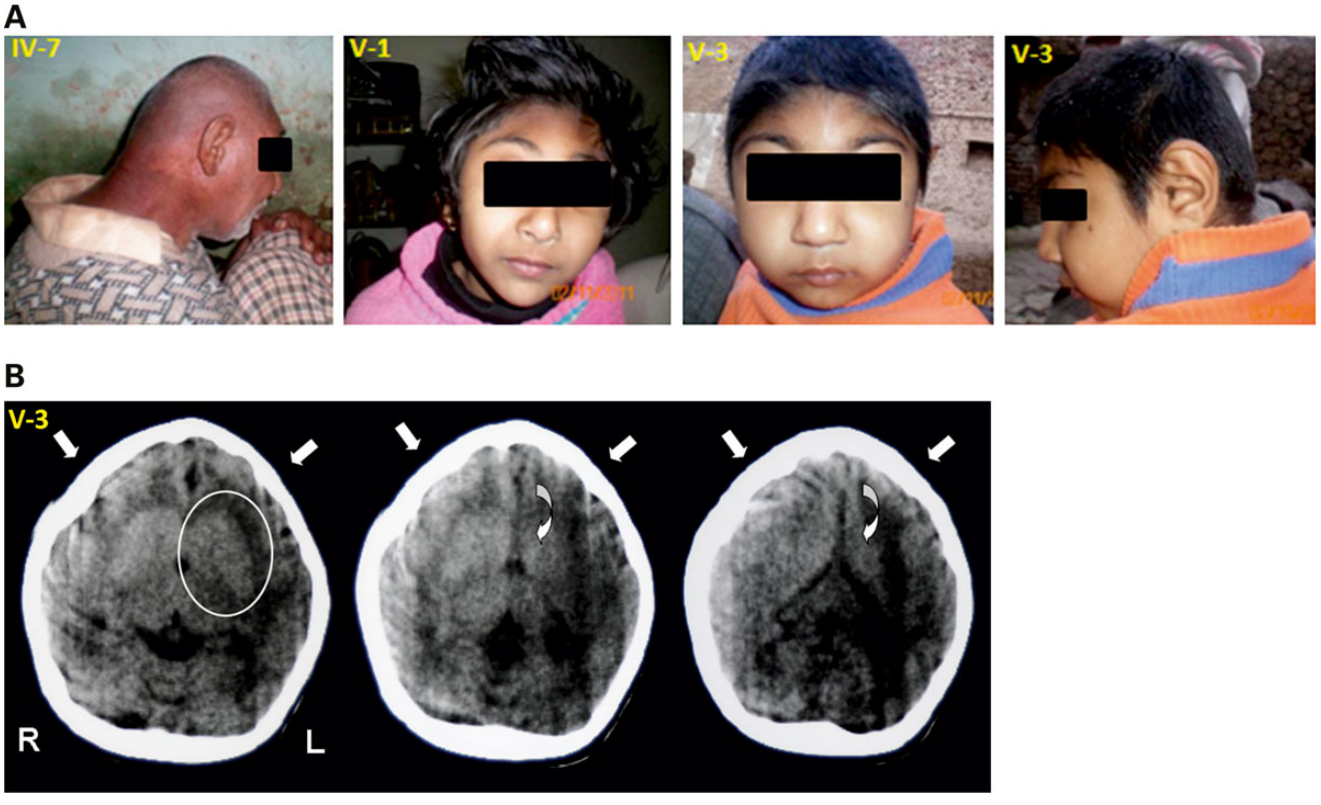

Figure 1. (**A**) Patients IV-7, V-1 and V-3. (**B**) Selected axial slices of cranial computed tomography obtained in a 3.5-year-old child (V-3) with autosomal recessive primary microcephaly. Notwithstanding compromised image quality, some distinct features can be noted: diminished cranial circumference with symmetrically poorly developed frontal lobes (arrows, also note deformed shape of the skull), poorly confined basal ganglia (oval) with missing delineation of the internal capsule, dysmorphic lateral ventricles (curved arrows, slit-like frontal horns) and infratentorial abnormality with hypoplasia of the vermis cerebelli (images not shown).

was listed in the NHLBI Exome Variant Server (http://eversusgs.washington.edu/EVS/), but only the mutation in *HsSAS-6* was predicted to have a potential pathogenic effect by the mutation prediction servers PolyPhen2 (HsSAS-6 p.Ile62Thr: score = 1.0; CAPZA1: p.Ser219Leu: score = 0.014) and ConDel (HsSAS-6 p.Ile62Thr: score = 0.847; CAPZA1: p.Ser219Leu: score = 0.001) (39,40). We further investigated whether the mutation in *CAPZA1* had any adverse effect on splicing by amplifying a 242-bp product from the cDNA that covered the junction between exons 8 and 9. This showed that splicing of *CAPZA1* is not affected by the mutation (Supplementary Material, Fig. S2). Moreover, previous knock-out experiments with this gene in the mouse failed to give evidence of a cell division phenotype (41). Overall, these results indicate that the mutation in *CAPZA1* is not causative of MCPH in the affected individuals.

Because of this, and because microcephaly is thought to result from defects in asymmetric cell division of neuronal progenitor cells, we concentrated on the *HsSAS-6* gene and its protein product, HsSAS-6, which is critical for the onset of centriole formation and thus for proper cell division (8,25,42–45). Human *SAS-6* consists of 17 exons and encodes a 657-amino acid protein of 74 kDa (46,47). The onset of centriole formation relies on the oligomerization of nine SAS-6 homodimers via their N-terminal head domains into a 9-fold symmetrical ring-like structure (30,48,49). Although the crystal structure of HsSAS-6 is not available, the high-resolution crystal structure of the N-terminal head domain of SAS-6 from *Danio rerio*, DrSAS-6, showed that Ile62 is part of the hydrophobic core of the protein (Fig. 4A). The side chain of this residue packs against those of Leu44, Leu60, Leu70, Phe80, Phe83 and Leu139, which are strictly conserved from mammals to algae (Supplementary Material, Fig. S3). This suggested that an exchange of the hydrophobic amino acid residue at position 62 with a polar threonine might compromise proper folding and/or function of the N-terminal domain of HsSAS-6.

In order to investigate the impact of the Ile62Thr mutation on HsSAS-6 function, we generated U2OS cells expressing doxycycline-inducible EGFP-tagged (hereafter referred to as GFP) HsSAS-6, either wild type (30) or bearing the Ile62Thr mutation. The cells were induced with doxycycline for 48 h, and the subcellular localization of the fusion proteins analyzed by immunofluorescence using antibodies against GFP and the centriolar marker Centrin-2. As shown in Figure 4B and C, we found that HsSAS-6 centriolar recruitment is not altered by the Ile62Thr mutation. These results demonstrate that the Ile62Thr mutation does not impair HsSAS-6 centriolar localization and that the protein is not misfolded.

Next, we analyzed whether the Ile62Thr mutation impairs the function of HsSAS-6 in centriole formation. Wild-type HsSAS-6 fused to GFP drives the formation of centrioles in excess (43), and we found that a variant bearing the Ile62Thr

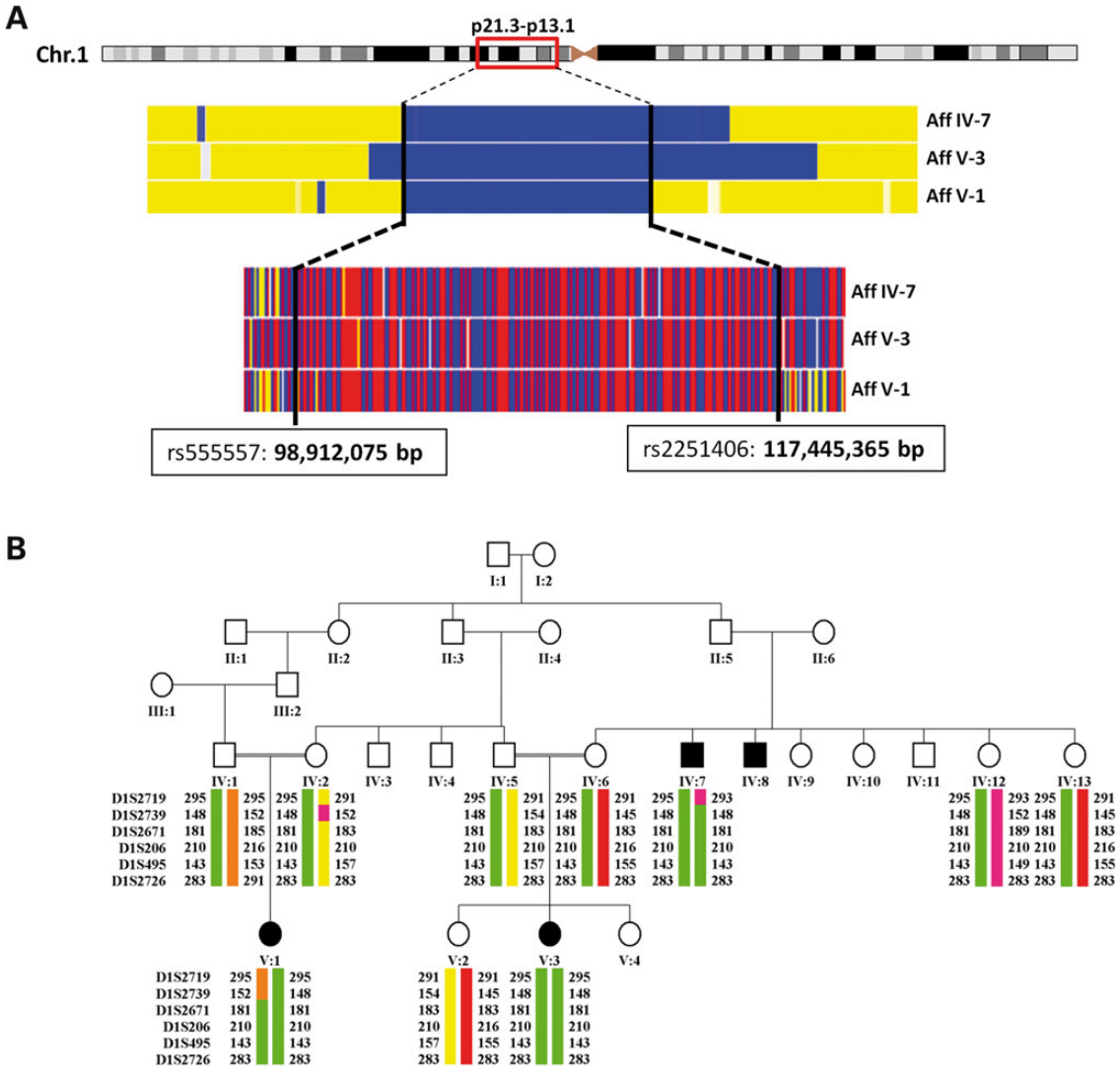

Figure 2. Homozygosity mapping. (**A**) 250 K NSP SNP array reveals a homozygous 18.53-Mb stretch (blue) with identical haplotype on Chromosome 1p21.3-p13.1 restricted by SNP rs555557 and rs2251406 in all three patients. (**B**) The green allele between STS marker D1S2671 and D1S2726 segregates in the family in an autosomal recessive manner.

Table 1. Clinical data of the affected

Patient pedigree ID	VI-7	VI-8	V-3	V-1
Age (years)	50	42	3.5	6
Gender	Male	Male	Female	Female
Occipitofrontal circumference	47 cm (SD = −6.63)	41.5 cm (SD = −10.26)	35 cm (SD = −19.6)	38.5 cm (SD = −15)
Height (cm)	155	157	82	95
Behavior	Aggressive	Aggressive	Aggressive	Aggressive
General physique	Weak	NORMAL	Weak	Normal
Epileptic shock	Yes	yes	No	No
Facial expression	Active	Active	Dull	Active
Attention	No	Low	Low	Low
Muscle tone	Normal	Normal	Normal	Normal
Adaptive behavior				
Self-feeding	Yes	Yes	Yes	Yes
Toilet training	No	Weak	No	No
Dressing/undressing training	No	No	No	No
Reading/writing	No	No	No	No
Self-care	No	No	Weak	No
Level of conveying their message	Very weak	Very weak	Very weak	Very weak

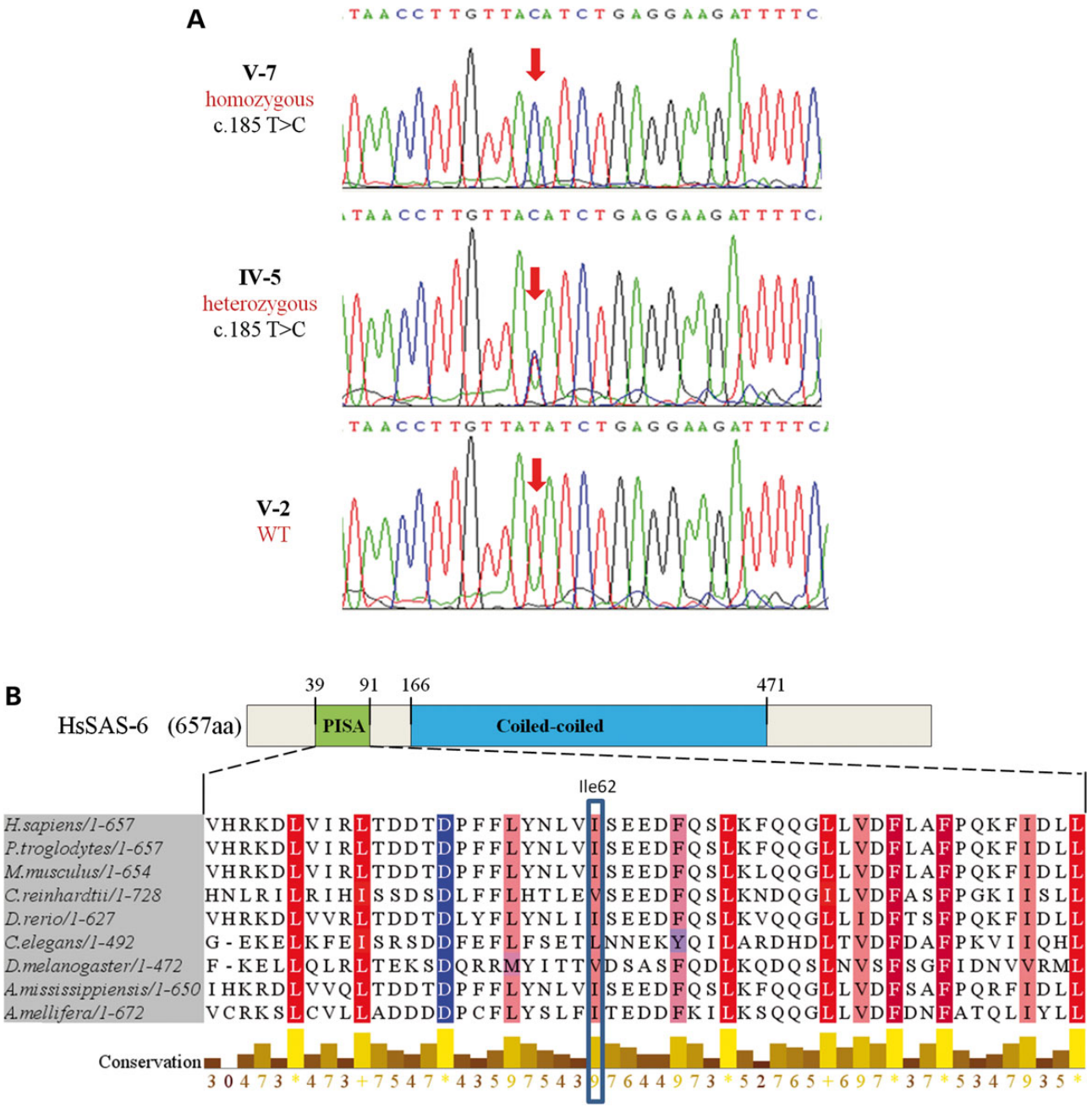

Figure 3. (**A**) Sanger sequencing of the affected patients revealed a homozygous c.185T>C transition. Non-affected family members were either heterozygous for the mutation (e.g. IV-5) or homozygous for the wild-type sequence. (**B**) The PISA domain of SAS-6 in 9 organisms was aligned in Jalview2.8 using the MAfftWS alignment and colored according to the hydrophobicity table of Kyte and Doolittle (37). Red indicates conserved hydrophobic and blue conserved hydrophilic residues. The shading intensity indicates the conservation grade of the hydrophobicity (dark = very conserved; white = not conserved). The conservation table shows that the properties of the mutated amino acid Ile62Thr in HsSAS-6 are highly conserved among different species (0 = no conservation; * or 11 = highest conservation, + or 10 = mutation but properties are conserved).

mutation does not (Fig. 4D and E). Furthermore, we found that ~7% of cells ($N = 186$) expressing the Ile62Thr variant exhibit less than three centrioles during mitosis, compared with ~1% for cells ($N = 229$) expressing wild-type HsSAS-6-GFP, indicative of a slight dominant-negative effect on centriole formation. Next, we set out to test whether the mutant variant can sustain centriole formation in cells depleted of endogenous HsSAS-6. To this end, we depleted endogenous HsSAS-6 using siRNAs directed against the 3′ UTR (30), which is absent from the GFP fusion constructs (Fig. 4D). We again analyzed cells by immunofluorescence with antibodies against GFP and centrin, but this time focusing on cells in mitosis to assay the number of centrioles at the end of the centriole duplication cycle. In control conditions, most cells had four centrioles, but ~12% of them harbored <4 centrin foci (Fig. 4F and H). In contrast, depletion of endogenous HsSAS-6 resulted

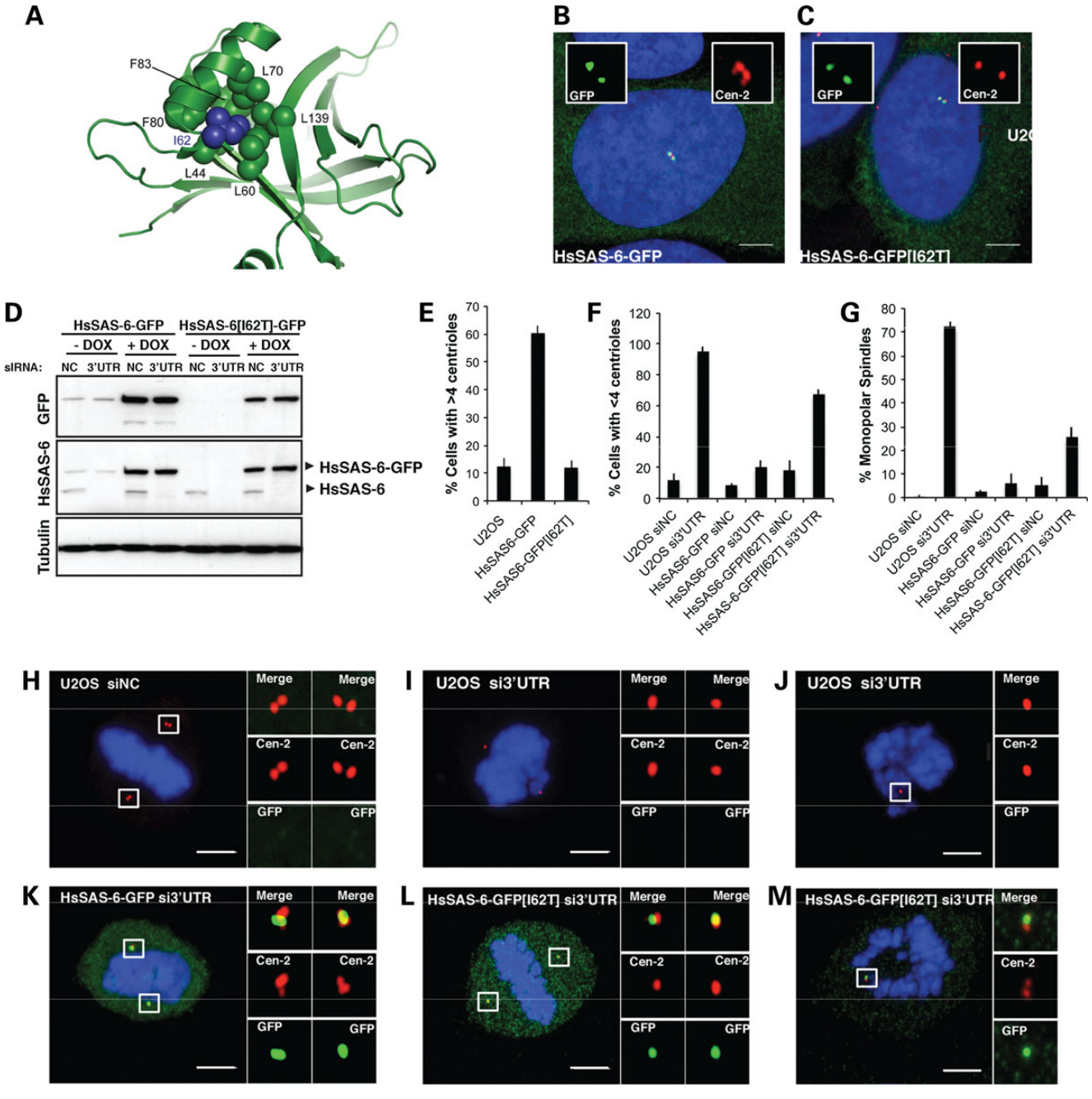

Figure 4. Impact of I62T mutation in tissue culture cells. (**A**) Cartoon representation of the N-terminal head domain of DrSAS-6 (PDB ID 2Y3V), illustrating the location and environment of Ile62. Selected amino acid residues are shown in sphere representation and are labeled. Ile62 is highlighted in blue. The figure was prepared using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.5. Schrödinger, LLC). (**B**–**C**) iU2OS interphase cells expressing either HsSAS-6 (wild type) fused to GFP (B) or the corresponding I62 T mutant (C) stained for GFP (green), Centrin-2 (red) and DNA (blue). (**D**) Western blot analysis of protein extracts from iU2OS cells expressing the indicated constructs and treated with control siRNAs (NC) or siRNAs targeting the 3′ UTR of HsSAS-6 (3′ UTR), using antibodies against GFP (top), HsSAS-6 (middle) or α-tubulin (bottom). Note that treatment with siRNAs against the HsSAS-6 3′ UTR leads to efficient depletion of endogenous HsSAS-6 but not of the GFP-variants that lack the 3′ UTR. (**E**) Histogram of the average frequency of mitotic cells with more than four centrioles in control U2OS cells and in cells overexpressing HsSAS-6-GFP or HsSAS-6-GFP[I62T] for 48 h following induction with doxycycline. Data from $\geq$3 experiments, with $>$50 cells scored in each experiment. (**F**–**G**) Histogram of the average frequencies of mitotic cells with less than four centrioles (F) or with monopolar spindles (G) in control U2OS, iU2OS HsSAS-6-GFP or iU2OS HsSAS-6-GFP[I62T] cells treated with siRNA Control (NC) or siRNA targeting the HsSAS-6 3′ UTR (3′ UTR). Data from $\geq$3 experiments, with $>$50 cells scored in each experiment. (**G**–**M**) U2OS (H–J), iU2OS HsSAS-6-GFP (K) or iU2OS HsSAS-6-GFP[I62T] (L–M) cells in mitosis, stained for GFP (green), Centrin-2 (red) and DNA (blue). In (L–M), doxycycline induction for 48 h was concomitant with siHsSAS-6 3′ UTR (3′ UTR) treatment. Note monopolar spindles in J and M. Scale bars: 5 μm.

in ~95% of cells having <4 centrin foci (Fig. 4F and I), with the majority of cell exhibiting monopolar spindle assembly (Fig. 4G and J). This centriole duplication phenotype was rescued to ~20% of cells having <4 centrin foci by the expression of wild-type HsSAS-6-GFP (Fig. 4 F and K). In contrast, ~68% of cells still exhibited <4 centrin foci upon the expression of HsSAS-6[Ile62Thr]-GFP (Fig. 4F and L). As anticipated also, monopolar spindle assembly was observed in a substantial fraction of cells expressing HsSAS-6[Ile62Thr]-GFP (Fig. 4G and M). We conclude that the Ile62Thr mutation severely, although not completely, impairs HsSAS-6 function. These data demonstrate that the c.185T>C mutation in the human *HsSAS-6* gene has a drastic impact on centriole formation and thus on normal cell division.

DISCUSSION

Microcephaly is thought to result from impaired asymmetric cell division of neuronal progenitors during the peak phase of neurogenesis in the embryo and is caused primarily by mutations in genes encoding centrosomal proteins (23), including some that are required for centriole formation (14,19,20,36,50). During the normal cell cycle, centriole formation is initiated around the G1/S transition, when a small set of centriolar proteins are recruited to the proximal end of the two parental centrioles, thus initiating assembly of one procentriole next to each parental centriole (49). Elongation continues throughout S phase and into G2, as well as into the next cell cycle, when the newly formed centrioles complete their maturation with the acquisition of appendages on their distal end (51–53).

As shown in this and in previous studies, knock-down of HsSAS-6 impairs procentriole formation, thus increasing the fraction of cells with fewer than four centrioles as well as cells with monopolar spindles (43,44,54). We have also shown that such impairment of procentriole formation can be rescued by overexpression of WT *HsSAS-6-GFP* but only to a limited extent by that of *HsSAS-6-Ile62Thr-GFP*, offering a strong cellular correlate to the phenotype manifested in the MCPH patients.

Our results are in line with a previous study on the Glu1235Val mutation in CPAP, which was one of the first mutations in this gene reported to cause microcephaly (14,24). CPAP is related to *C. elegans* SAS-4 and is also essential for procentriole formation in proliferating human cells in culture (55). CPAP localizes to the proximal part of the procentriole and centriole and has been proposed to be connected with HsSAS-6 through the bridging protein CEP135 (35). Mirroring our findings with the Ile62Thr HsSAS-6 mutant, the Glu1235Val CPAP mutant can rescue centriole formation only to a limited extent (24). These observations taken together offer a striking parallel and reinforce the notion that partial impairment of centriole formation results in MCPH.

Furthermore, studies in *Drosophila* larvae showed that compared with wild-type flies, *DmSas-6* knockout flies exhibit a significantly reduced number of centrosomes in the brain and ~18% of centrosomes were even smaller in size (56).

In conclusion, we have demonstrated here that the homozygous c.185T>C mutation in the *HsSAS-6* gene has a drastic impact on centriole formation and thus on proper cell division, a process that is essential during neurogenesis. Furthermore, we propose that the remaining activity of the mutated HsSAS-6[I62T] protein enables residual asymmetric cell division and thus results in reduced brain development, causing primary autosomal recessive microcephaly.

MATERIAL AND METHODS

Sample collection

After obtaining the informed consent, blood was drawn from three affected and seven healthy family members from the Pakistani family investigated in this study and genomic DNA was isolated according to standard protocols. For cDNA analysis, blood from V-3 and VI-7 as well as from two healthy controls was collected in PAXgene RNA blood tubes (PreAnalytiX). RNA was further isolated using PAXgene RNA Kit. Moreover, blood was drawn from 116 non-affected Pakistani individuals for control analysis. The study was approved by Institutional ethical review boards of Gomal University, Dera Ismail Khan, and Quaid-i-Azam University, Islamabad, Pakistan.

Autozygosity mapping, haplotype and linkage analysis

All three affected individuals, IV-7, V-1 and V-3, were genotyped on Affymetrix GeneChip Human Mapping 250K NspI Arrays at the 'Center for Medical Research' at the Medical University of Graz.

Genome-wide linkage analysis of the family was performed with 20 044 selected SNP markers. LOD scores were calculated with ALLEGRO (57). Data handling, evaluation and statistical analysis were performed as described previously (19).

For haplotype and local linkage analysis, a total of six highly polymorphic STR markers covering the autozygous region of interest were selected for fine mapping and segregation analysis, including D1S206 (101.6 Mb) and D1S2726 (111.18 Mb) from the ABI Prism Linkage Mapping Set v2.5, as well as D1S2719 (96.81 Mb), D1S2739 (98.93 Mb), D1S2671 (101.67 Mb) and D1S495 (102.56 Mb) selected from the UCSC browser mapping track (build 37/ hg19) (47). PCRs were performed with ABI Prism True Allele PCR Premix (Applied Biosystems), and amplicons were denaturated using HiDiFormamide with Gene Ruler 500-Liz Size Standard (both from Applied Biosystems). Genome scan data were generated on the ABI3130*xl* and analyzed with Peak Scanner Software v1.0 (Applied Biosystems). For linkage analysis, an autosomal recessive trait with full penetrance and a disease allele frequency of 0.001 were assumed. The two-point LOD score was calculated using the online version of Superlink (http://bioinfo.cs.technion.ac.il/superlink-online), and for multipoint LOD score calculation, Merlin was used (58). Sex-averaged recombination rates between markers were obtained from Rutgers map (build 37, patch 4) (59).

Whole-exome sequencing

We fragmented 1 μg of DNA using sonification technology (Covaris, Woburn, MA, USA). The fragments were end-repaired and adaptor-ligated including incorporation of sample index barcodes. After size selection, the library was subjected to the enrichment process. For that we chose the SeqCap EZ Human

Exome Library v2.0 kit from NimbleGen (Roche NimbleGen, Madison, WI, USA). The enriched library was subsequently sequenced on an Illumina HiSeq 2000 sequencing instrument using a paired-end 2 × 100-bp protocol.

This resulted in 8.4 Gb of mapped sequences, a mean coverage of 89-fold, a 30× coverage of 87% and a 10× coverage of 97% of target sequences. For data analysis, the Varbank pipeline v.2.3 and filter interface was used (unpublished, https://varbank.ccg.uni-koeln.de/). Primary data were filtered according to signal purity by the Illumina Realtime Analysis (RTA) software v1.8. Subsequently, the reads were mapped to the human genome reference build hg19 using the BWA (60) alignment algorithm. GATK v1.6 (61) was used to mark duplicated reads, to do a local realignment around short insertion and deletions, to recalibrate the base quality scores and to call SNPs and short indels.

Scripts developed in-house at the Cologne Center for Genomics were applied to detect protein changes, affected donor and acceptor splice sites and overlaps with known variants. Acceptor and donor splice site mutations were analyzed with a Maximum Entropy model (62) and filtered for effect changes. In particular, we filtered for high-quality (coverage >15; quality >25) rare (MAF < 0.005) homozygous variants (dbSNP build 135, the 1000 Genomes database build 20110521, and the public Exome Variant Server, NHLBI Exome Sequencing Project, Seattle, build ESP6500). We also filtered against an in-house database containing variants from 511 exomes from epilepsy patients to exclude pipeline-related artifacts (MAF < 0.004).

Sanger sequencing

Primers covering the coding exons and splice sites *of HsSAS-6*, *NGF*, *PSCR1* and *WDR47* were designed with Primer3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) (63). A list of all oligonucleotides can be found in Supplementary Material, Table S2. CAPZA1_cDNA_forw 5′ GGAAGTTCAC CATCACACCA and CAPZA1_cDNA_rev 5′ GGCCTTGAAT GTGGTATCTGA primers were used to investigate the effect of the splice site mutation in exon 8 of CAPZA1 on mRNA level. PCR was performed using HotStarTaq Master Mix Kit (Qiagen) with the following cycling conditions: 94°C for 15 min, followed by 35 cycles of 95°C for 25 s, 57°C for 30 s and 72°C for 1 min with no final elongation step. The Sanger sequencing reaction was set up with the Big Dye v3.1 cycling sequencing kit (Applied Biosystems) according to the manufacturers' protocol. Dye remnants were removed with Centri-Sep™ columns (Applied Biosystems). Bidirectional DNA sequencing was conducted on the ABI3130xl (Applied Biosystems), and the data were analyzed with ChromasLite software (Technelysium Pty Ltd.) and the UCSC browser (47,64).

HsSAS-6 homology analysis

HsSAS-6 I62T homology among nine different species [from NCBI, *Homo sapiens* (AAI01027), *D. rerio* (AAI65167), *Pan troglodytes* (JAA36473), *Aligator mississippienis* (XP_006277337), *Apis melifera* (XP_395972), *Mus musculus* (NP_082625), *C. elegans* (CAA16384), *Drosophila melanogaster* (AAF56983) and *Chlamydomonas reinhardtii* (BAF94334)] was investigated with the desktop version of Jalview2.8 (65). Sequences were aligned with MafftWS alignment using the default settings (66). Regions that were highly conserved in terms of their hydrophobicity are highlighted in red and blue.

Plasmids for human cell experiments

The pEBTet-GFP plasmids (67) were obtained from Dirk Gründemann. The following oligos were annealed GW-F (5′ CG CGGGTACCGCCGGCAGCTAGCGGCGCGCCCGGCCGA TAT), GW-R (5′ ATATCGGCCGGGCGCGCCGCTAGCTGC CGGCGGTACCCGCG), digested with KpnI and EagI and ligated into KpnI, NotI cut pEBTet-GFP producing the pEBTet-MCS vector. This plasmid was then used to insert fluorescence proteins and Gateway cassette (Invitrogen), generating the destination vector pEBTet-GW-EGFP. The multiple cloning site of pENTR 1A (Invitrogen) was modified by introducing single restriction sites between the attR1 and attR2 sites (3′ AgeI and XbaI 5′), generating the entry vector pENTR-SD-Age-AGT. Full-length HsSAS-6 (NM_194292.1) was amplified using the primers Age-Ko-HsSAS6-F (5′ CGCGACCGGTAC CATGAGCCAAGTGCTGTTCCAC) and Xba-noST-S6-R (5′ CGCGTCTAG ATAACTGTTTGGTAACTGCCCA) and cloned into pENTR-SD-Age vector by restriction digest with AgeI and XbaI.

Mutations of the I62 residue in HsSAS-6 were generated by site-directed mutagenesis on pENTR-SD-Age-HsSAS-6 using the following primers: S6-I62T-fwd, *ACA*TCTGAGGAA GATTTTCAAAGT and S6-I62T-rev, AACAAGGTTATA TAAAAAAAATGG (mutated codon is italicized).

Gateway reaction was then performed according to the manufacturer's protocol to generate the expression plasmid pEBTet-HsSAS-6-GFP and pEBTet-HsSAS-6-I62T, which were sequence verified.

Cell culture and transfections

U2OS cells were obtained from the ECACC and maintained in McCoy's 5A GlutaMAX medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS) for U2OS cells or tetracycline-negative FBS (Brunschwig) for the inducible episomal cell lines (iU2OS). iU2OS cell lines were generated by transfecting U2OS cells with pEBTet-HsSAS-6-GFP or pEBTet-HsSAS-6[I62T]-GFP using Lipofectamine2000 (Invitrogen). Transfected cells were selected with 1 μg/ml of puromycin 1 day after transfection and amplified. For the expression of the GFP-fused constructs, early passage cells were induced with 1 μg/ml of doxycycline for 48 h.

Endogenous HsSAS-6 was depleted using a Stealth RNAi™ siRNA (Invitrogen) targeting the 3′ UTR of HsSAS-6 (5′ GAG CUGUUAAAGACUGGAUACUUUA 3′) (30). Stealth RNAi™ siRNA-negative control LO GC (Invitrogen) was used as a control.

siRNA transfection was performed using LipofectamineRNAiMax (Invitrogen) according to the manufacturer's protocol, and cells were analyzed 48–72 h after siRNA treatment.

Cell-extract preparation and biochemical assays

Cells were collected, washed in PBS and lysed on ice for 1 h in lysis buffer [50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.5 mM

EDTA 0.5% NP-40, Complete Mini Protease Inhibitor Cocktail (Roche Diagnostics)]. Lysates were cleared by centrifugation for 10 min at 12 000 × g and 4°C before the supernatant was collected. SDS–PAGE was performed using 10% polyacrylamide gels (BioRad), followed by transfer on nitrocellulose membrane (Amersham). The membrane was probed with mouse HsSAS-6 antibody (Santa Cruz, 1:1000) or mouse α-tubulin antibody (Sigma, 1:10000), followed by incubation with HRP-conjugated secondary (Promega) and the signal detected with Chemiluminescence (Roche).

Immunofluorescence and microscopy for human cells

U2OS cells grown on glass coverslips were fixed for 7–10 min in −20°C methanol, washed in PBS and blocked for 15–30 min in 1% bovine serum albumin and 0.05% Tween-20 in PBS. Cells were incubated 2 h at room temperature with primary antibodies, washed three times for 10 min in PBST (0.05% Tween-20 in PBS) incubated for 45 min at room temperature with secondary antibodies, stained with ∼1 μg/ml of Hoechst 33258, washed three times in PBST and mounted. Primary antibodies were 1 : 4000 mouse centrin (20H5; gift from Jeffrey L. Salisbury) and 1 : 500 rabbit GFP (gift from Viesturs Simanis). Secondary antibodies were 1 : 1000 goat anti-rabbit coupled to Alexa 488 and 1 : 1000 goat anti-mouse coupled to Alexa 568. For quantification of centrioles, mitotic cells (prophase to metaphase) with similar cytoplasmic GFP expression were used; highly expressing cells that often harbored GFP aggregates were not retained for analysis. Imaging was done on a Zeiss LSM710 confocal microscope. Optical sections were acquired every 0.12 μm, and planes containing centrioles were projected together. Images were processed using ImageJ and Adobe Photoshop, preserving relative image intensities within a series.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are thankful to the family members for their willingness to participate with this study.

Conflict of Interest statement. None declared.

FUNDING

M.O. received funding from the Federation of European Biochemical Societies and the European Research Council (AdG 233335 to P.G.).

REFERENCES

1. Verloes, A., Drunat, S., Gressens, P. and Passemard, S. (2009–2013) Primary autosomal recessive microcephalies and seckel syndrome spectrum disorders. In Pagon, R.A., Adam, M.P., Ardinger, H.H., Bird, T.D., Dolan, C.R., Fong, C.T., Smith, R.J.H. and Stephens, K. (eds.), *GeneReviews [Internet]*. University of Washington, Seattle, WA, available from: http://www.ncbi.nlm.nih.gov/books/NBK9587/.
2. Mochida, G.H. and Walsh, C.A. (2001) Molecular genetics of human microcephaly. *Curr. Opin. Neurol.*, **14**, 151–156.
3. Darvish, H., Esmaeeli-Nieh, S., Monajemi, G.B., Mohseni, M., Ghasemi-Firouzabadi, S., Abedini, S.S., Bahman, I., Jamali, P., Azimi, S., Mojahedi, F. *et al.* (2010) A clinical and molecular genetic study of 112 Iranian families with primary microcephaly. *J. Med. Genet.*, **47**, 823–828.
4. Desir, J., Cassart, M., David, P., Van Bogaert, P. and Abramowicz, M. (2008) Primary microcephaly with ASPM mutation shows simplified cortical gyration with antero-posterior gradient pre- and post-natally. *Am. J. Med. Genet. A.*, **146A**, 1439–1443.
5. Cox, J., Jackson, A.P., Bond, J. and Woods, C.G. (2006) What primary microcephaly can tell us about brain growth. *Trends Mol. Med.*, **12**, 358–366.
6. Mahmood, S., Ahmad, W. and Hassan, M.J. (2011) Autosomal recessive primary microcephaly (MCPH): clinical manifestations, genetic heterogeneity and mutation continuum. *Orphanet J. Rare Dis.*, **6**, 39.
7. Woods, C.G., Bond, J. and Enard, W. (2005) Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am. J. Hum. Genet.*, **76**, 717–728.
8. Thornton, G.K. and Woods, C.G. (2009) Primary microcephaly: do all roads lead to Rome? *Trends Genet.*, **25**, 501–510.
9. Sajid Hussain, M., Marriam Bakhtiar, S., Farooq, M., Anjum, I., Janzen, E., Reza Toliat, M., Eiberg, H., Kjaer, K.W., Tommerup, N., Noegel, A.A. *et al.* (2013) Genetic heterogeneity in Pakistani microcephaly families. *Clin. Genet.*, **83**, 446–451.
10. O'Driscoll, M., Jackson, A.P. and Jeggo, P.A. (2006) Microcephalin: a causal link between impaired damage response signalling and microcephaly. *Cell Cycle*, **5**, 2339–2344.
11. Bilguvar, K., Ozturk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatli, B., Yalnizoglu, D., Tuysuz, B., Caglayan, A.O., Gokben, S. *et al.* (2010) Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*, **467**, 207–210.
12. Nicholas, A.K., Khurshid, M., Desir, J., Carvalho, O.P., Cox, J.J., Thornton, G., Kausar, R., Ansar, M., Ahmad, W., Verloes, A. *et al.* (2010) WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat. Genet.*, **42**, 1010–1014.
13. Yu, T.W., Mochida, G.H., Tischfield, D.J., Sgaier, S.K., Flores-Sarnat, L., Sergi, C.M., Topcu, M., McDonald, M.T., Barry, B.J., Felie, J.M. *et al.* (2010) Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat. Genet.*, **42**, 1015–1020.
14. Bond, J., Roberts, E., Springell, K., Lizarraga, S.B., Scott, S., Higgins, J., Hampshire, D.J., Morrison, E.E., Leal, G.F., Silva, E.O. *et al.* (2005) A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat. Genet.*, **37**, 353–355.
15. Genin, A., Desir, J., Lambert, N., Biervliet, M., Van Der Aa, N., Pierquin, G., Killian, A., Tosi, M., Urbina, M., Lefort, A. *et al.* (2012) Kinetochore KMN network gene CASC5 mutated in primary microcephaly. *Hum. Mol. Genet.*, **21**, 5306–5317.
16. Bond, J., Roberts, E., Mochida, G.H., Hampshire, D.J., Scott, S., Askham, J.M., Springell, K., Mahadevan, M., Crow, Y.J., Markham, A.F. *et al.* (2002) ASPM is a major determinant of cerebral cortical size. *Nat. Genet.*, **32**, 316–320.
17. Kumar, A., Girimaji, S.C., Duvvari, M.R. and Blanton, S.H. (2009) Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am. J. Hum. Genet.*, **84**, 286–290.
18. Sir, J.H., Barr, A.R., Nicholas, A.K., Carvalho, O.P., Khurshid, M., Sossick, A., Reichelt, S., D'Santos, C., Woods, C.G. and Gergely, F. (2011) A primary microcephaly protein complex forms a ring around parental centrioles. *Nat. Genet.*, **43**, 1147–1153.
19. Hussain, M.S., Baig, S.M., Neumann, S., Peche, V.S., Szczepanski, S., Nurnberg, G., Tariq, M., Jameel, M., Khan, T.N., Fatima, A. *et al.* (2013) CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. *Hum. Mol. Genet.*, **22**, 5199–5214.
20. Hussain, M.S., Baig, S.M., Neumann, S., Nurnberg, G., Farooq, M., Ahmad, I., Alef, T., Hennies, H.C., Technau, M., Altmuller, J. *et al.* (2012) A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. *Am. J. Hum. Genet.*, **90**, 871–878.
21. Awad, S., Al-Dosari, M.S., Al-Yacoub, N., Colak, D., Salih, M.A., Alkuraya, F.S. and Poizat, C. (2013) Mutation in PHC1 implicates chromatin remodeling in primary microcephaly pathogenesis. *Hum. Mol. Genet.*, **22**, 2200–2213.
22. Yang, Y.J., Baltus, A.E., Mathew, R.S., Murphy, E.A., Evrony, G.D., Gonzalez, D.M., Wang, E.P., Marshall-Walker, C.A., Barry, B.J., Murn, J. *et al.* (2012) Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation and differentiation. *Cell*, **151**, 1097–1112.

23. Wang, X., Tsai, J.W., Imai, J.H., Lian, W.N., Vallee, R.B. and Shi, S.H. (2009) Asymmetric centrosome inheritance maintains neural progenitors in the neocortex. *Nature*, **461**, 947–955.
24. Kitagawa, D., Kohlmaier, G., Keller, D., Strnad, P., Balestra, F.R., Fluckiger, I. and Gönczy, P. (2011) Spindle positioning in human cells relies on proper centriole formation and on the microcephaly proteins CPAP and, STIL. *J. Cell. Sci.*, **124**, 3884–3893.
25. Wollnik, B. (2010) A common mechanism for microcephaly. *Nat. Genet.*, **42**, 923–924.
26. Lander, E.S. and Botstein, D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, **236**, 1567–1570.
27. Alkuraya, F.S. (2010) Homozygosity mapping: one more tool in the clinical geneticist's toolbox. *Genet. Med.*, **12**, 236–239.
28. Alkuraya, F.S. (2013) The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum. Genet.*, **132**, 1197–1211.
29. Hildebrandt, F., Heeringa, S.F., Ruschendorf, F., Attanasio, M., Nurnberg, G., Becker, C., Seelow, D., Huebner, N., Chernin, G., Vlangos, C.N. *et al.* (2009) A systematic approach to mapping recessive disease genes in individuals from outbred populations. *PLoS Genet.*, **5**, e1000353.
30. Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M.C., Fluckiger, I., Gönczy, P. *et al.* (2011) Structural basis of the 9-fold symmetry of centrioles. *Cell*, **144**, 364–375.
31. van Breugel, M., Hirono, M., Andreeva, A., Yanagisawa, H.A., Yamaguchi, S., Nakazawa, Y., Morgner, N., Petrovich, M., Ebong, I.O., Robinson, C.V. *et al.* (2011) Structures of SAS-6 suggest its organization in centrioles. *Science*, **331**, 1196–1199.
32. American Psychiatric Association. (2013) *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Publishing, Arlington, VA.
33. Guernsey, D.L., Jiang, H., Hussin, J., Arnold, M., Bouyakdan, K., Perry, S., Babineau-Sturk, T., Beis, J., Dumas, N., Evans, S.C. *et al.* (2010) Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. *Am. J. Hum. Genet.*, **87**, 40–51.
34. Tang, C.J., Lin, S.Y., Hsu, W.B., Lin, Y.N., Wu, C.T., Lin, Y.C., Chang, C.W., Wu, K.S. and Tang, T.K. (2011) The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. *EMBO J.*, **30**, 4790–4804.
35. Lin, Y.C., Chang, C.W., Hsu, W.B., Tang, C.J., Lin, Y.N., Chou, E.J., Wu, C.T. and Tang, T.K. (2013) Human microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly. *EMBO J.*, **32**, 1141–1154.
36. Brown, N.J., Marjanovic, M., Luders, J., Stracker, T.H. and Costanzo, V. (2013) Cep63 and cep152 cooperate to ensure centriole duplication. *PLoS One*, **8**, e69986.
37. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
38. Maun, N.A., Speicher, D.W., DiNubile, M.J. and Southwick, F.S. (1996) Purification and properties of a ca(2+)-independent barbed-end actin filament capping protein, CapZ, from human polymorphonuclear leukocytes. *Biochemistry*, **35**, 3518–3524.
39. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
40. Gonzalez-Perez, A. and Lopez-Bigas, N. (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am. J. Hum. Genet.*, **88**, 440–449.
41. Neumann, B., Walter, T., Heriche, J.K., Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I., Held, M., Liebel, U. *et al.* (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, **464**, 721–727.
42. Woods, C.G. and Parker, A. (2013) Investigating microcephaly. *Arch. Dis. Child.*, **98**, 707–713.
43. Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A. and Gönczy, P. (2007) Regulated HsSAS-6 levels ensure formation of a single procentriole per centriole during the centrosome duplication cycle. *Dev. Cell.*, **13**, 203–213.
44. Leidel, S., Delattre, M., Cerutti, L., Baumer, K. and Gönczy, P. (2005) SAS-6 defines a protein family required for centrosome duplication in, *C. elegans* and in human cells. *Nat. Cell Biol.*, **7**, 115–125.
45. Nakazawa, Y., Hiraki, M., Kamiya, R. and Hirono, M. (2007) SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Curr. Biol.*, **17**, 2169–2174.
46. UniProt Consortium. (2012) Reorganizing the protein space at the universal protein resource (UniProt). *Nucl. Acids Res.*, **40**, D71–D75.
47. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at, UCSC. *Genome Res.*, **12**, 996–1006.
48. Hilbert, M., Erat, M.C., Hachet, V., Guichard, P., Blank, I.D., Fluckiger, I., Slater, L., Lowe, E.D., Hatzopoulos, G.N., Steinmetz, M.O. *et al.* (2013) *Caenorhabditis elegans* centriolar protein SAS-6 forms a spiral that is consistent with imparting a ninefold symmetry. *Proc. Natl. Acad. Sci. USA*, **110**, 11373–11378.
49. Gönczy, P. (2012) Towards a molecular architecture of centriole assembly. *Nat. Rev. Mol. Cell Biol.*, **13**, 425–435.
50. Cottee, M.A., Muschalik, N., Wong, Y.L., Johnson, C.M., Johnson, S., Andreeva, A., Oegema, K., Lea, S.M., Raff, J.W. and van Breugel, M. (2013) Crystal structures of the CPAP/STIL complex reveal its role in centriole assembly and human microcephaly. *Elife*, **2**, e01071.
51. Vorobjev, I.A. and Chentsov, Y. (1982) Centrioles in the cell cycle. *I. Epithelial Cells J. Cell Biol.*, **93**, 938–949.
52. Kuriyama, R. and Borisy, G.G. (1981) Centriole cycle in Chinese hamster ovary cells as determined by whole-mount electron microscopy. *J. Cell Biol.*, **91**, 814–821.
53. Chretien, D., Buendia, B., Fuller, S.D. and Karsenti, E. (1997) Reconstruction of the centrosome cycle from cryoelectron micrographs. *J. Struct. Biol.*, **120**, 117–133.
54. Dammermann, A., Muller-Reichert, T., Pelletier, L., Habermann, B., Desai, A. and Oegema, K. (2004) Centriole assembly requires both centriolar and pericentriolar material proteins. *Dev. Cell.*, **7**, 815–829.
55. Kohlmaier, G., Loncarek, J., Meng, X., McEwen, B.F., Mogensen, M.M., Spektor, A., Dynlacht, B.D., Khodjakov, A. and Gönczy, P. (2009) Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. *Curr. Biol.*, **19**, 1012–1018.
56. Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G. and Glover, D.M. (2007) DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. *Curr. Biol.*, **17**, 1465–1472.
57. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L. and Kong, A. (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.*, **25**, 12–13.
58. Abecasis, G.R., Cherny, S.S., Cookson, W.O. and Cardon, L.R. (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.*, **30**, 97–101.
59. Matise, T.C., Chen, F., Chen, W., De La Vega, F.M., Hansen, M., He, C., Hyland, F.C., Kennedy, G.C., Kong, X., Murray, S.S. *et al.* (2007) A second-generation combined linkage physical map of the human genome. *Genome Res.*, **17**, 1783–1786.
60. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.
61. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
62. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
63. Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, **132**, 365–386.
64. Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A. *et al.* (2011) The UCSC genome browser database: Update 2011. *Nucl. Acids Res.*, **39**, D876–D882.
65. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
66. Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucl. Acids Res.*, **33**, 511–518.
67. Bach, M., Grigat, S., Pawlik, B., Fork, C., Utermohlen, O., Pal, S., Banczyk, D., Lazar, A., Schomig, E. and Grundemann, D. (2007) Fast set-up of doxycycline-inducible protein expression in human cell lines with a single plasmid based on epstein-barr virus replication and the simple tetracycline repressor. *FEBS J.*, **274**, 783–790.