nature genetics

Mutations in *CEP57* cause mosaic variegated aneuploidy syndrome

Katie Snape[1,16], Sandra Hanks[1,16], Elise Ruark[1], Patricio Barros-Núñez[2], Anna Elliott[1], Anne Murray[1], Andrew H Lane[3], Nora Shannon[4], Patrick Callier[5], David Chitayat[6], Jill Clayton-Smith[7], David R FitzPatrick[8], David Gisselsson[9], Sebastien Jacquemont[10], Keiko Asakura-Hay[11], Mark A Micale[12], John Tolmie[13], Peter D Turnpenny[14], Michael Wright[15], Jenny Douglas[1] & Nazneen Rahman[1]

Using exome sequencing and a variant prioritization strategy that focuses on loss-of-function variants, we identified biallelic, loss-of-function *CEP57* mutations as a cause of constitutional mosaic aneuploidies. CEP57 is a centrosomal protein and is involved in nucleating and stabilizing microtubules. Our findings indicate that these and/or additional functions of CEP57 are crucial for maintaining correct chromosomal number during cell division.

Cell division is a highly complex process that involves chromosomal duplication and separation into two daughter cells. Errors in this process result in gains or losses of chromosomes, which is known as aneuploidy. Aneuploidy is an important cause of human disease that is causally implicated in miscarriage, developmental disorders and cancer[1,2]. Multiple biological processes, including mitotic checkpoints, chromatid attachment mechanisms and centrosome, kinetochore and microtubule functions, are required to prevent aneuploidy[1,2].

Mosaic variegated aneuploidy syndrome (MVA; MIM257300) is a rare autosomal recessive disorder characterized by mosaic aneuploidies, diverse phenotypic abnormalities and predisposition to cancer[3,4]. Study of this rare condition can inform on the basic processes controlling aneuploidy in humans. For example, we previously showed that biallelic *BUB1B* mutations cause MVA, linking mitotic spindle defects to human disease and providing some of the strongest evidence that aneuploidy can be a cause, as well as a consequence, of cancer[4].

Because *BUB1B* mutations underlie only a proportion of MVA, we adopted an exome sequencing strategy to identify additional causes of this disease. We undertook exome sequencing in two siblings with MVA (individuals with ID numbers 633_1 and 633_2) in whom we had previously excluded *BUB1B* mutations[3]. We used exome capture followed by paired-end sequencing on an Illumina GAIIx Solexa sequencer (**Supplementary Methods**). Ninety percent of target bases in the resulting sequence had ≥10× coverage. We used NextGENe software (see URLs) to detect sequence variants and applied a series of filters to identify variants that were most likely to be causative for MVA (**Supplementary Methods** and **Supplementary Table 1**). We removed variants with read coverage <10 and/or with a wild-type to mutant ratio of <30%, as these are more likely to be false. We also removed variants in known pseudogenes. As MVA is a very rare condition, we removed all variants present in 50 exomes we have previously sequenced in individuals with other conditions and all variants recorded in the dbSNP and/or 1000 Genomes database. We also removed all synonymous and intronic variants outside splice junctions, as these are unlikely to be causative. After applying these filters, there were 439 variants in sibling 1 and 439 variants in sibling 2. We used this final dataset in the analyses outlined below.

We first applied a loss-of-function (LOF) script that identifies nonsense mutations, coding insertions or deletions that result in translational frameshifts and insertions, deletions or base substitutions at consensus splice residues (**Supplementary Methods**). We removed LOF variants in genes in which we had identified >1 LOF variant in the 50 in-house exomes on the assumption that LOF variants detected in such genes are too common to cause MVA and/or are more likely to be artifactual. After this, there remained 20 LOF variants in sibling 1 and 24 variants in sibling 2 (**Supplementary Table 2**). We initially focused our analyses on LOF variants because of the strong prior evidence of association of this class of mutation with human disease.

There was no history of consanguinity in family 633, and therefore our expectation was for the siblings to be compound heterozygotes for two different pathogenic mutations in the same gene. We identified no genes with two different LOF mutations and 12 genes with one LOF mutation shared by the two siblings (**Supplementary Table 2**). We next applied a conventional recessive script to identify all genes with two variants present in both siblings. There were no genes with one LOF mutation together with a second variant. There were four genes with two non-LOF variants that were both present in each sibling, but none were likely to be causative (**Supplementary Table 3**).

[1]Section of Cancer Genetics, Institute of Cancer Research, Sutton, UK. [2]División de Genética, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, México. [3]Division of Pediatric Endocrinology, State University of New York at Stony Brook, Stony Brook, New York, USA. [4]Clinical Genetics Service, City Hospital, Nottingham, UK. [5]Service de Cytogénétique, Centre Hospitalier Universitaire (CHU) de Dijon, Dijon, France. [6]The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital and Division of Clinical & Metabolic Genetics, Department of Pediatrics, Hospital for Sick Children, Toronto, Ontario, Canada. [7]Manchester Biomedical Research Center, Manchester Academic Health Sciences Center, St. Mary's Hospital, Manchester, UK. [8]Medical Research Council (MRC) Human Genetics Unit, Institute of Genetic and Molecular Medicine, Western General Hospital, Edinburgh, UK. [9]Department of Clinical Genetics, Lund University, University and Regional Laboratories, Skåne University Hospital, Lund, Sweden. [10]Division of Medical Genetics, CHU Vaudois, Lausanne, Switzerland. [11]Ninewells Hospital and Medical School, Tayside University Hospitals National Health Service (NHS) Trust, Dundee, UK. [12]Department of Anatomic Pathology, William Beaumont Hospital, Royal Oak & Oakland University William Beaumont School of Medicine, Rochester, Michigan, USA. [13]Department of Clinical Genetics, Ferguson-Smith Centre, Royal Hospital for Sick Children, Glasgow, UK. [14]Peninsula Clinical Genetics Service and Peninsula Medical School, Royal Devon & Exeter Healthcare NHS Trust, Exeter, UK. [15]Northern Genetics Service, Newcastle upon Tyne Hospitals Foundation Trust, Newcastle upon Tyne, UK. [16]These authors contributed equally to this work. Correspondence should be addressed to N.R. (nazneen.rahman@icr.ac.uk).

Received 22 December 2010; accepted 6 April 2011; published online 8 May 2011; doi:10.1038/ng.822

Table 1 ***CEP57*** **mutations and associated clinical features**

Case ID		663_1	663_2	638	657
Mutations		c.520_521delGA c.915_925dup11	c.520_521delGA c.915_925dup11	c.241C>T; p.R81X c.241C>T; p.R81X	c.915_925dup11 c.915_925dup11
Age		8.5 yrs	4.5 yrs	Died 3 wks	Died 15 yrs
Clinical features[a]	Mosaic aneuploidies	Yes	Yes	Yes	Yes
	Growth retardation	Yes	Yes	Yes	Yes
	Microcephaly	No	Yes	No	Yes
	Intrauterine growth retardation	No	Yes	Yes	No
	Mental retardation	No	No	n/a	Mild
	Congenital heart disease	No	No	Yes	Mild
	Hypothyroidism	Yes	No	n/a	Yes
	Rhizomelic shortening	No	No	Yes	Yes
	Other features	None	None	Duodenal atresia; hypotonia	Hearing impairment; sleep apnea
Reference		3	3	4	7

[a]More detailed clinical information is given in the **Supplementary Methods**.

Of the 12 genes with one shared LOF mutation in the siblings, *CEP57* was notable because it localizes to the centrosome[5,6]. Using Sanger sequencing, we confirmed the *CEP57* c.520_521delGA deletion detected through the exome sequencing and identified a second LOF mutation, an 11-bp insertion, c.915_925dup11, which was also present in both siblings (**Supplementary Methods**, **Supplementary Figs. 1** and **2** and **Supplementary Table 4**). This insertion is a duplication of the preceding 11 bases (**Supplementary Fig. 2**). Analysis of parental DNA showed that the deletion was present in the father, and the duplication was present in the mother, consistent with autosomal recessive inheritance. On review of the exome data, the 11-bp insertion was clearly present in 30–40% of reads in each sibling. However, this insertion had not been called in many reads, particularly those in which the insertion was toward the end of the read, presumably because the reduced flanking sequence impaired alignment. As a result, the mutant to wild-type read ratio was below the NextGENe calling threshold in both siblings and the insertion was initially disregarded (**Supplementary Table 1** and **Supplementary Fig. 3**).

We next used Sanger sequencing to analyze *CEP57* in 18 affected individuals from 13 additional *BUB1B*-negative families with MVA and identified two further individuals with biallelic *CEP57* mutations, confirming the causative role of *CEP57* in aneuploidy predisposition (**Table 1** and **Supplementary Fig. 2**). Child 638 was homozygous for a nonsense mutation (c.241C>T; p.Arg81X), and each of his parents, who are first cousins, are heterozygous for the mutation. Child 657 is homozygous for c.915_925dup11, and using dosage analysis, we confirmed that the duplication was present on both of his alleles (**Supplementary Methods**, **Supplementary Fig. 4** and **Supplementary Table 5**). This child was fostered, and we could not obtain parental samples or information regarding a possible relationship between his parents[7].

There were some shared clinical features in the individuals with *CEP57* mutations (**Table 1** and **Supplementary Methods**). All the affected individuals had random gains and losses of chromosomes in ~25–50% of examined cells (**Fig. 1a**). In addition, they all had growth retardation at the milder end of the spectrum but with relative sparing of the head. There was no gross dysmorphology, and their development was normal or mildly delayed. Of note, two of the affected individuals had hypothyroidism and two had rhizomelic shortening of the upper limbs, which may be specific associated features of *CEP57* mutations.

CEP57 encodes a 500-amino-acid protein[5]. Secondary structure prediction suggests it is composed of two α-helical coiled-coil domains connected by a flexible linker region (**Fig. 1b**). The N-terminal coiled-coil domain is within a region required for localization of CEP57 to the centrosome and for multimerization of the protein. The C-terminal half of CEP57, including the second coiled-coil domain, is required for nucleating, bundling and anchoring microtubules to the centrosomes within basket-like structures[6]. In addition, CEP57 is involved in intracellular bidirectional trafficking of factors such as FGF2 along microtubules[8].

The centrosomal localization and microtubular stabilization functions may explain why CEP57 loss results in aneuploidy. However, it should be noted that CEP57 is a relatively understudied protein, and it may be that other, currently uncharacterized functions result in the phenotype we observed in individuals with biallelic inactivating *CEP57* mutations. *BUB1B*, which encodes BUBR1, primarily prevents aneuploidy through a mechanism in which CEP57 is currently not known to have a role, functioning in the mitotic spindle checkpoint and to maintain stable interactions between microtubules and the kinetochore[9,10]. The phenotypes associated with *BUB1B* and *CEP57* mutations are broadly similar, although *BUB1B* mutations are strongly associated with cancer[3,4]. Thus far, no cancers have been reported in the small number of *CEP57* mutation-positive individuals identified, although none have reached adulthood. The role of *CEP57*

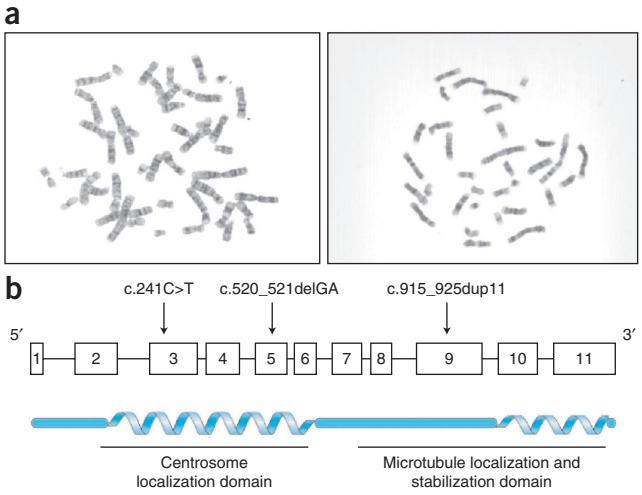

Figure 1 *CEP57* structure, mutations and resulting aneuploidy. (**a**) Aneuploid metaphase karyotypes from family 633 showing gains and losses of whole chromosomes. (**b**) Schematic representation of the genomic structure of *CEP57* with the positions of the mutations indicated with arrows. Below is the protein structure showing the coiled-coil domains, the centrosomal localization region and the microtubular stabilization region.

as a somatic target in cancer has also not been well interrogated[11]. Given that aneuploidy is present in >90% of solid tumors, our results suggest that further analyses would be of interest.

Exome analysis is proving successful in identifying many Mendelian disease genes[12]. However, the downstream processes of variant calling, filtering, interpretation and prioritization for follow up remain challenging. In particular, many commonly used algorithms have found insertion and deletion variants difficult to call with high sensitivity and specificity[13]. As a result, thus far, the majority of new disease genes identified through exome analyses were discovered through the detection of base-substitution mutations[12]. Although some LOF mutations, such as nonsense and some splicing mutations, result from base substitutions, the majority are caused by insertions or deletions. Thus, currently, many exome analyses are not optimally configured for disease gene identification. We have optimized a calling strategy with good sensitivity for small insertions and deletions, and this allowed us to detect one of the *CEP57* mutations (a 2-bp deletion). However, large insertions, such as the second *CEP57* mutation, remain challenging to call, and particular focus on improving insertion and deletion detection is required for the full potential of exomic analyses to be realized.

URLs. NextGENe, http://www.softgenetics.com/NextGENe.html; UCSC Genome Bioinformatics, http://genome.ucsc.edu/; Ensembl, http://www.ensembl.org/index.html; NCBI SNP database, http://www.ncbi.nlm.nih.gov/projects/SNP/; 1000 Genomes, http://www.1000genomes.org/; HUGO Gene Nomenclature Database, http://www.genenames.org/.

Accession codes. Human *CEP57* cDNA sequence is available in GenBank under the accession number NM_014679, and human CEP57 protein sequence is available in UniProtKB under the accession number Q86XR8.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We are very grateful to all the families that participated in this study and to A. Zachariou, M. Warren-Perry, R. Linger, D. Dudakia and J. Bull for assistance in their recruitment. We thank D. Leongamornlert for assistance with exome sequencing, A. Renwick, S. Seal, G. Bowden, D. Hughes and C. Turnbull for assistance in Solexa optimization and validation experiments and B. Ebbs for DNA extraction and for running the ABI sequencer. We thank K. Tatton-Brown for assistance with clinical review of cases, A.I. Vasquez and J. Swansbury for the karyotyping and A. Strydom for secretarial assistance with the manuscript. K.S. is supported by the Michael and Betty Kadoorie Cancer Genetics Research Programme. We acknowledge NHS funding to the National Institute for Health Research Biomedical Research Centre. This work was supported by Cancer Research UK (grants C8620_A9024 and C8620_A8857) and the Institute of Cancer Research (UK).

AUTHOR CONTRIBUTIONS

S.H., J.D. and A.M. performed exome sequencing. K.S., E.R., A.E. and N.R. performed data management and analysis. S.H. and J.D. performed *CEP57* mutation and dosage analysis. P.B.-N., A.H.L., N.S., P.C., D.C., J.C.-S., D.R.F., D.G., S.J., K.A.-H., M.A.M., J.T., P.D.T. and M.W. provided clinical material. N.R. wrote the manuscript with substantial input from K.S., S.H., E.R. and J.D. N.R. designed and oversaw all aspects of the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Holland, A.J. *et al. Nat. Rev. Mol. Cell Biol.* **10**, 478–487 (2009).
2. Torres, E.M. *et al. Genetics* **179**, 737–746 (2008).
3. García-Castillo, H. *et al. Am. J. Med. Genet. A.* **146A**, 1687–1695 (2008).
4. Hanks, S. *et al. Nat. Genet.* **36**, 1159–1161 (2004).
5. Andersen, J.S. *et al. Nature* **426**, 570–574 (2003).
6. Momotani, K. *et al. Biochem. J.* **412**, 265–273 (2008).
7. Lane, A.H. *et al. Am. J. Med. Genet.* **110**, 273–277 (2002).
8. Meunier, S. *et al. Traffic* **10**, 1765–1772 (2009).
9. Musacchio, A. *et al. Nat. Rev. Mol. Cell Biol.* **8**, 379–393 (2007).
10. Suijkerbuijk, S.J. *et al. Cancer Res.* **70**, 4891–4900 (2010).
11. Forbes, S.A. *et al. Nucleic Acids Res.* **38**, D652–D657 (2010).
12. Ng, S.B. *et al. Hum. Mol. Genet.* **19**, R119–R124 (2010).
13. Ng, S.B. *et al. Nat. Genet.* **42**, 790–793 (2010).