Hypomorphism in human NSMCE2 linked to primordial dwarfism and insulin resistance

Felicity Payne, Rita Colnaghi, Nuno Rocha, Asha Seth, Julie Harris, Gillian Carpenter, William E. Bottomley, Eleanor Wheeler, Stephen Wong, Vladimir Saudek, David Savage, Stephen O’Rahilly, Jean-Claude Carel, Inês Barroso, Mark O’Driscoll, and Robert Semple

Introduction

Primordial dwarfism denotes extreme linear growth impairment from early gestation. When severely reduced head size is also seen, primordial dwarfism is said to be microcephalic. This is generally taken to imply a global problem with either cellular or organisational growth and contrasts with primordial dwarfism in which the head size is relatively preserved, which is more likely to be accounted for by a more tissue-selective problem, affecting, for example, long bone growth (1).

Many human genetic defects have now been described as underlying microcephalic primordial dwarfism. These may affect endocrine growth regulation by impairing production or action of IGF-1 (2, 3); however, most mutations reported to date affect cell-autonomous growth. Some clustering of defects according to the cellular structure or process affected is apparent. For example, microcephalic primordial dwarfism can result from defects in genes involved in the licensing of DNA replication, in centrosome function, or in DNA damage response and repair pathways (reviewed in ref. 4).

In keeping with the known roles of the defective genes, some syndromes feature a characteristic profile of chromosomal instability or impaired DNA damage repair. Some components of the syndromes cannot easily be explained by known functions of the affected gene, however. An example is the severe insulin resistance described in pericentrin deficiency (5). This usually appears during the first decade of life and is associated with severely fatty liver, elevated blood triglyceride levels, and diabetes, recapitulating in severe form the pandemic metabolic syndrome. Similar metabolic abnormalities are seen with milder growth impairment in both Bloom syndrome (6) and Werner syndrome (7), each caused by loss of a DNA helicase, and without growth impairment in patients with a heterozygous polymerase-inactivating mutation in DNA polymerase δ (8). However despite increasing evidence of a link between impaired DNA replication and/or damage responses and severe insulin resistance, the underlying mechanisms linking these remain unclear.

Results

A novel syndrome of primordial dwarfism. We identified 2 female patients with similar syndromes (details in Supplemental Material and Methods; supplemental material available online with this article; doi:10.1172/JCI73264DS1). The shared features are (a) severe primordial dwarfism, with facial dysmorphism, including
were identified; however, 3 genes, \textit{NSMCE2}, \textit{RUSC1}, and \textit{WDR36}, harbored compound heterozygous variants. Detailed analysis of these genes in large human populations, and prior functional studies, suggested only the \textit{NSMCE2} variants to be plausible candidates to cause the observed phenotype (details in Supplemental Appendix).

\textit{NSMCE2} contained compound heterozygous frameshift mutations: c.345delT (p.Ser116Leufs*18), inherited from the proband’s mother, and c.700_701insAGGG (p.Ala234Glufs*4), inherited from the patient’s father (Figure 1H). Both mutations were absent from 1,092 individuals from the 1000 Genomes Project Consortium (9) and 4,190 internal control exomes and genomes. The p.Ala234Glufs*4 was also absent from 6,250 individuals from the NHLBI GO Exome Sequencing Project, while the p.Ser116Leufs*18 mutation was found in heterozygous form in 2 of the 6,250 individuals (allele frequency 0.00016) (details in Methods and Supplemental Table 2).

P2 was the only child of unrelated parents of European ancestry from the Brittany region of northwest France. Sanger sequencing of \textit{NSMCE2} in P2 revealed the same 2 frameshift mutations found in P1. Her mother was heterozygous only for the p.Ala234Glufs*4 mutation, while her father was heterozygous for the p.Ser116Leufs*18 mutation, confirming P2 to be a compound heterozygote (Figure 1L). Both mutations were absent from 1,092 individuals from the 1000 Genomes Project Consortium (9) and 4,190 internal control exomes and genomes. The p.Ala234Glufs*4 was also absent from 6,250 individuals from the NHLBI GO Exome Sequencing Project, while the p.Ser116Leufs*18 mutation was found in heterozygous form in 2 of the 6,250 individuals (allele frequency 0.00016) (details in Methods and Supplemental Table 2).

P2 was the only child of unrelated parents of European ancestry from the Brittany region of northwest France. Sanger sequencing of \textit{NSMCE2} in P2 revealed the same 2 frameshift mutations found in P1. Her mother was heterozygous only for the p.Ala234Glufs*4 mutation, while her father was heterozygous for the p.Ser116Leufs*18 mutation, confirming P2 to be a compound heterozygote (Figure 1L). Although the families of P1 and P2 have no known shared ancestry, genotyping of 21 SNPs and 2 microsatellites within the linkage...
The NSMCE2/MMS21 gene product is often named MMS21 due to its original identification in a Saccharomyces cerevisiae screen for mutants sensitive to the alkylating agent methane methylsulphonate (MMS) (13). The Schizosaccharomyces pombe gene is Nse2. Here, we refer to the protein as NSMCE2/MMS21, or NSMCE2 in figures.

The function of SMC5–6, although critical for survival in all organisms studied, is not well understood. Current evidence, particularly from hypomorphic yeast mutants, has defined roles in DNA replication and recombination, including the facilitation of homologous recombination repair (HRR), through aiding of resolution of topologically complex DNA intermediates. Failure to complete HRR may lead to checkpoint activation and apoptosis or to catastrophic failure to untangle highly catenated, interconnected chromosomes.

NSMCE2/MMS21 possesses a Siz/PIAS RING (SP-RING) domain toward its C terminus with SUMO E3-ligase activity. Several block encompassing NSMCE2 (chr8: 126104083–1263936) confirmed that both patients inherited the same 2 rare haplotypes within a region greater than 510 kb, one containing the p.Ser116Leufs*18 mutation, the other p.Ala234Glufs*4. This suggests shared common ancestral haplotypes: one, greater than 510-kb long, containing p.Ser116Leufs*18, estimated to have arisen 283 generations (less than 8,500 years) ago, and the other, greater than 339-kb long, containing p.Ala234Glufs*4, estimated to have arisen 150 generations (less than 4,500 years) ago, based on the locus-specific recombination rate. Neither haplotype nor the p.Ser116Leufs*18 nor p.Ala234Glufs*4 mutations were present in a sample of 54 controls from Great Britain (Methods and Supplemental Table 3).

NSMCE2/MMS21 and SMC5–6. NSMCE2 encodes a protein component of the heteromultimeric structural maintenance of chromosomes 5–6 (SMC5–6) complex (10). It consists of 2 core SMC components (SMC5 and SMC6) and, in humans, 4 non-SMC elements often designated as NSE1-4, among which NSE2 corresponds to NSMCE2/MMS21 (10–12). The NSMCE2 gene product is often named MMS21 due to its original identification in a Saccharomyces cerevisiae screen for mutants sensitive to the alkylating agent methane methylsulphonate (MMS) (13). The Schizosaccharomyces pombe gene is Nse2. Here, we refer to the protein as NSMCE2/MMS21, or NSMCE2 in figures.

The function of SMC5–6, although critical for survival in all organisms studied, is not well understood. Current evidence, particularly from hypomorphic yeast mutants, has defined roles in DNA replication and recombination, including the facilitation of homologous recombination repair (HRR), through aiding of resolution of topologically complex DNA intermediates. Failure to complete HRR may lead to checkpoint activation and apoptosis or to catastrophic failure to untangle highly catenated, interconnected chromosomes.

NSMCE2/MMS21 possesses a Siz/PIAS RING (SP-RING) domain toward its C terminus with SUMO E3-ligase activity. Sev-
mRNA levels of NSMCE2/MMS21 were 50% of WT in dermal fibroblasts, although near normal in LCLs. cDNA sequencing showed no evidence of the p.Ser116Leufs*18 species, suggesting that its mRNA is unstable (Supplemental Figure 2, C and D). We did not detect truncated NSMCE2/MMS21 protein. However on heterologous overexpression of a Myc-tagged version of NSMCE2/MMS21 carrying the p.Ser116Leufs*18 mutation in A549 cells, a low level of the truncated protein could be seen on immunoblotting (Supplemental Figure 3A). Although p.Ala234Glufs*4, in contrast, was detected at levels close to those of overexpressed WT NSMCE2/MMS21, its levels decayed more quickly than those of WT NSMCE2/MMS21 after overexpression in HeLa cells treated with cycloheximide, consistent with reduced protein stability (Supplemental Figure 2E).

To determine the impact of the mutations on the SUMO-ligase activity of NSMCE2/MMS21, we assessed auto-SUMOylation of Myc-tagged NSMCE2/MMS21 ectopically coexpressed with GFP-SUMO1 in A549 cells. NSMCE2/MMS21 species were immunoprecipitated using anti-Myc beads, and the SUMO-modified versions of NSMCE2/MMS21 were detected by Western blotting with antibodies specific to human NSMCE2/MMS21 (Supplemental Figure 3). Typical micrographs of binucleated dermal fibroblasts showing (A and B) MN or (C) NPB. Scale bars: 5 μm. (D) Increased frequency of MN or (E) NPB in Cyt-B–treated dermal fibroblasts from P1 relative to age-matched healthy controls (WT) following a 4-hour block in 1 mM HU. (F) DOX-induced expression of WT or SUMO LD NSMCE2/MMS21 in P1 or control dermal fibroblasts (WT fibroblasts, 2 samples shown) analyzed by immunoblotting with antibodies specific to human NSMCE2/MMS21 (WB: α-NSMCE2). Asterisk indicates a crossreactive band that serves as an internal loading control. (G and H) DOX-induced ectopic expression of WT NSMCE2/MMS21 reduces frequencies of MN (G) and NPB (H) to levels seen in control cells. In contrast, DOX-induced ectopic expression of NSMCE2/MMS21 (LD) significantly increases levels of both MN (G) and NPB (H). Frequencies of MN per 100 or NPB per 1,000 binucleated cells are shown for each experimental condition, as indicated. All bar graphs represent mean ± SEM (n = 4). *P < 0.05; **P < 0.01; ***P < 0.001 (unpaired 1-tailed t test).
p. Ala234Glufs*4 mutant exhibited a level of auto-SUMOylation similar to that of WT NSMCE2/MMS21 under these conditions (Figure 2, E and F).

Having confirmed that cells from P1 were severely hypomorphic for NSMCE2/MMS21, we next characterized the associated cellular phenotype. The frequencies of micronuclei (MN) and nucleoplasmic bridges (NPB) in cytokinesis-blocked binucleated cells (Figure 3, A–C) are widely used as indicators of chromosomal aberrations. A previously validated SUMO-ligase-deficient (LD) version of NSMCE2/MMS21 with mutations in the SP-RING (p.Cys185Ala and p.His187Ala) (19) was used as a negative control. Consistent with its reduced expression and lack of SP-RING domain, auto–SUMO-ligase activity of p.Ser116Leufs*18 was virtually undetectable. As expected, auto-SUMOylation activity of the LD variant was also severely reduced. In contrast, the auto–SUMO-ligase activity of p.Ala234Glufs*4 mutant exhibited a level of auto-SUMOylation similar to that of WT NSMCE2/MMS21 under these conditions (Figure 2, E and F).

Having confirmed that cells from P1 were severely hypomorphic for NSMCE2/MMS21, we next characterized the associated cellular phenotype. The frequencies of micronuclei (MN) and nucleoplasmic bridges (NPB) in cytokinesis-blocked binucleated cells (Figure 3, A–C) are widely used as indicators of chromosomal aberrations.
damage and instability (20). Typically, MN develop when errors in chromosome segregation during mitosis occur from acentric fragments of chromosomes and/or anaphase-lagging whole chromosomes, while NPB form when centromeres of dicentric chromosomes are drawn to opposite poles at anaphase or by telomere shortening. When the actin polymerization inhibitor cytochalasin B (Cyt-B) is used to inhibit cytokinesis, observed MN are derived from events in the preceding S phase. Given the role suggested for SMC5–6 in resolution of HRR intermediates, we thus assessed the frequencies of MN and NPB in Cyt-B–induced binucleated dermal fibroblasts from P1.

The basal frequency of MN was significantly higher in binucleated dermal fibroblasts from P1 relative to WT, although NPB frequencies in these cells were indistinguishable (Figure 3, D and E). We next assessed these following treatment with hydroxyurea (HU), a potent inducer of replication fork stalling and collapse. Following HU, binucleated dermal fibroblasts from P1 showed higher frequencies of both MN and NPB than WT (Figure 3, D and E). Doxycycline-induced (DOX-induced) expression of WT NSMCE2/MMS21 (Figure 3F) significantly reduced the frequency of MN and NPB in P1 binucleates. In contrast, expression of the LD version of NSMCE2/MMS21 markedly increased their frequency (Figure 3, G and H).

LCLs derived from P1 exhibited a similar phenotype. There was an approximately 1.8-fold increase in spontaneously binucleated cells compared with WT, an increase also observed in Bloom syndrome LCLs. This increase was exacerbated in both P1 and Bloom syndrome LCLs following exposure to repeated treatment with low concentrations of HU (once daily for 4 days) (Figure 4A). Reminiscent of P1 fibroblasts, we also observed modestly elevated spontaneous and HU-induced MN formation in binucleates generated by Cyt-B treatment (Figure 4B).

Yeast SUMO-LD mutants of the NSMCE2/MMS21 ortholog and mutations in smc6 are associated with replication delay and accumulation of unresolved hemi-catenated, X-shaped DNA molecules (21, 22). BrdU pulse chase analysis of asynchronously growing P1-derived LCLs did not reveal a significant difference in S phase progression rate compared with WT (Supplemental Fig-
Bloom syndrome helicase (BLM), which is disrupted in Bloom syndrome, plays a fundamental role in resolution of converging Holliday junctions during HRR. Consistent with synthetic interaction between the BLM ortholog in yeast (Sgs1) and mutation of the NSMCE2/MMS21 ortholog, we observed an impairment in BLM foci formation in P1-derived LCLs both in response to replication stress induced by HU and in response to ionizing radiation (IR) (Figure 4E), in contrast with WT. A similar impairment of BLM focus formation was seen in patient fibroblasts exposed to HU (Figure 4F). Moreover, following UV treatment, another potent means of replication fork stalling, we observed significant elevation in UV-induced sister chromatid exchanges (SCE) in P1 LCLs compared with WT, consistent with the impaired ability to form BLM foci of these cells (Figure 4G). Collectively, these data are consistent with a failure of NSMCE2/MMS21-mutated patient cells to complete cell division effectively following chronic repli- cative stress. In contrast, P1 LCLs showed no hypersensitivity to camptothecin or mechlorethamine, unlike BLM-deficient cells (Supplemental Figure 5, A–C). As HRR-defective cells are sensitive to killing by both these agents, this suggests that DNA repair per se is not completely deficient as a result of this hypomorphic defect in NSMCE2/MMS2.

Modeling of NSMCE2/MMS21 deficiency in zebrafish. As no Nsmce2 knockout mouse has been reported to date, we used zebrafish to test the consequence of loss of Nsmce2 function in vivo. We utilized a cocktail of 2 splice-site anti-sense morpholinos (MOs) to deplete the zebrafish NSMCE2/MMS21 ortholog, reducing its mRNA levels in the morphant to approximately 20% of those of control MO-injected embryos (Methods and Supplemental Figure 6). This markedly reduced body size compared with uninjected and control MO-injected embryos (Figure 5, A and B). From 5% to 15% of the nsmce2/mms21 MO-injected embryos exhibited an extreme phenotype including body curvature and increased mortality; however, the majority of embryos were grossly normal morphologically. Coinjection of human WT NSMCE2 mRNA with zebrafish nsmce2/mms21-targeting MOs attenuated the dwarf phenotype observed in the nsmce2 morphant alone, with coinjected embryos not significantly different in length from those injected with control MO alone (Figure 5, C–E). In contrast, coinjection of either mRNA variant identified in the patients failed to rescue the nsmce2/mms21 morphant phenotype (Figure 5, C and D).

In keeping with findings in mammalian cells, levels of GFP-tagged mutant NSMCE2/MMS21 protein decayed much more rapidly than those of GFP-tagged WT NSMCE2/MMS21, despite high and equivalent levels of mRNA being detected after injection, suggesting impaired translation and/or instability of the mutants (Supplemental Figure 7). Collectively, these experiments confirm the specificity of the MO effect and suggest that the NSMCE2/MMS21 frameshift mutations identified in the patients lead to loss of function. Interestingly, coinjection of a SUMO LD version of NSMCE2/MMS21 also failed to rescue nsmce2 morphants (Figure 5E) coincident with reduced expression of this GFP-tagged protein, as seen for the other mutants (Supplemental Figure 7).

Discussion

Our findings demonstrate association between compound heterozygosity for rare frameshift mutations in NSMCE2/MMS21 and a syndrome of primordial dwarfism, extreme insulin resistance, and primary gonadal failure. One of the mutations, p.Ser116Leufs*18, removes the whole SP-RING domain, is nearly undetectable in primary cells, has undetectable SUMO autoligase activity, and is unable to rescue the zebrafish morphant phenotype on reexpression. The second mutation, p.Ala234Glufs*4, in contrast, is predicted to remove only 14 amino acids distal to the SP-RING domain, is well expressed from a strong promoter in heterolo- gous cellular systems, and exhibits normal SUMO autoligation after overexpression. The direct evidence that it is a loss-of-function mutation thus rests on the extremely low level of full-length NSMCE2/MMS21 observed in primary fibroblasts and LCLs from P1 and on the failure of the reexpressed mutant to rescue the dwarfism of nsmce2/mms21 morphant zebrafish, which we suggest is due to reduced stability of the mutant protein. These data establish that the patients described are severely hypomorphic for NSMCE2/MMS21 and demonstrate that this results in severe dwarfism, insulin resistance and primary gonadal failure, with mild evidence of spontaneous chromosomal instability.

The SMC5–6 complex is structurally closely related to cohesin (SMC1–3) and condensin (SMC2–4). Although cohesin and condensin have relatively well characterized roles in mitosis, respectively, holding sister chromatids in apposition or mediating the premitotic condensation of chromosomes (23), the function of the SMC5–6 complex is less well understood. The accumulation of aberrant DNA intermediates with hypofunction of the yeast NSMCE2/MMS21 ortholog (21, 22) and synthetic lethality in combination with mutations in the topoisomerase II ortholog (24) argues for a role in resolving topologically complex intermediates of HRR. Inability to resolve these intermediates is particularly harmful in the context of restarting stalled or collapsed DNA replication forks, which appear, for example, when the DNA template is damaged or the nucleotide pool depleted (21, 22, 25).

The phenotype of patient cells, namely impaired ability to recover from replication stress leading to genomic instability both at base- line and in response to induced replication fork stalling, is in keeping with a role for NSMCE2/MMS21 in resolving complex DNA structures and is consistent with findings in yeast, avian, and mammalian cells (e.g., refs. 11, 14–16, 26). We also demonstrate a mild increase in SCE following UV irradiation of patient LCLs; however, this increase is much smaller than that seen in Bloom syndrome. Such an aberrant “hyperrecombination” phenotype has also been reported in chick DT40 cells with NSMCE2/MMS21 deficiency (26).

The precise role of the SUMO-ligase activity of NSMCE2/MMS21 is not fully delineated. Several targets have been documented in various organisms (14–16). As well as NSMCE2/MMS21, SMC5, and SMC6, these include other proteins with important roles in DNA damage sensing and chromosome stability, including Ku70, Scc1, and the telomeric proteins TRF1 and TRF2. Studies in avian (26) and mammalian (16, 27) cells have suggested that loss of the SUMO-ligase activity of NSMCE2/MMS21 orthologs leads to genomic instability. Our data now extend these findings. The increased MN and NPB seen after reexpression of a SUMO LD variant of NSMCE2/MMS21 in patient fibroblasts suggest that
SUMO-ligase activity may be required for at least some important functions of NSMCE2/MMS21 in higher eukaryotes. Nevertheless, the viability of yeast strains lacking ligase activity and the nonviability of strains wholly lacking the NSMCE2/MMS21 ortholog indicate that NSMCE2/MMS21 has other essential functions that are SUMO-ligase independent.

The question arises as to whether the cellular defects we observe account for the phenotype of the affected patients. Growing cells are subjected to a barrage of endogenously generated DNA damage during every cell division, arising, for example, from generation of reactive oxygen and nitrogen species as a by-product of metabolism (28), so it is plausible that reduced cellular ability to tolerate replicative stress and associated DNA damage is the cause of global growth impairment. This is mechanistically akin to the primordial dwarfism seen when the DNA damage-sensing kinase ATR (29) is mutated in Seckel syndrome.

Both patients with NSMCE2/MMS21 deficiency also exhibit primary ovarian failure. Primary gonadal failure is a consequence of defects in several other human DNA damage response components, including ATM, NBS, BLM and members of the Fanconi anemia complex, presumably due to impaired meiosis and/or mitosis of germ cells in utero. Furthermore, the yeast ortholog of NSMCE2/MMS21 has been shown to be required for meiosis (30). The ovarian failure of the two patients we describe is thus potentially explicable through established functions of NSMCE2/MMS21. It remains possible, however, that NSMCE2/MMS21 serves unknown functions that are related to the observed phenotype, as in the case of genetic cohesinopathies, where important components of the clinical syndromes result from perturbed transcriptional regulation, rather than directly to loss of cohesin’s longest established role in sister chromatid cohesion.

The highly insulin resistant diabetes that we report is more difficult to reconcile simply with the observed cellular defects. As our study ascertained patients based on the presence of severe insulin resistance, further work will be needed to determine confidently whether it is an obligate feature of NSMCE2/MMS21 loss of function in humans; however the insulin resistance in both cases was extreme. Although the derangement in insulin and glucose levels was of a similar degree to that seen with loss of insulin receptor function (31), the metabolic phenotype was quite distinct. Both our patients had severe fatty liver and dyslipidemia, which are not seen with insulin receptor dysfunction (32). Instead, the metabolic phenotype closely resembles that seen in primary disorders of adipose tissue development or function (33).

A clue to the origins of the metabolic derangement we report may come from its resemblance to that of patients with Bloom and Werner syndromes (6, 7) or of patients with heterozygous polymerase domain mutations in DNA polymerase δ (8). Bloom and Werner syndromes are caused, respectively, by loss of the BLM (RECQL3) and WRN (RECQL2) DNA helicases, both of which play multiple accessory roles in resolution of complex DNA structures. The yeast phenotype caused by loss of the primordial ancestral DNA helicase that later diverged to produce WRN and BLM resembles the phenotype seen with hypomorphism for the NSMCE2/MMS21 ortholog (22). Our observation of impaired replicative stress-induced BLM foci formation and modestly increased levels of SCE following UV irradiation in NSMCE2/MMS21 patient cells further argues for commonality between Bloom syndrome and the syndrome we describe. On the other hand, our patients did not have cardinal clinical features of either Bloom or Werner syndrome, and the level of increase of SCE was far lower here than in Bloom syndrome. The detailed cellular phenotype of cells with polymerase domain mutations of DNA polymerase δ, the predominant lagging strand DNA polymerase, has yet to be reported.

Although it is tempting to equate whole-body insulin sensitivity to cell-autonomous insulin sensitivity, this is commonly misleading. Maintenance of whole-body insulin sensitivity requires not only intact cellular insulin signaling pathways, but also normal size and function of key insulin-responsive tissues — particularly adipose tissue, muscle, and liver. Indeed, most Mendelian forms of insulin resistance are accounted for by mutations, not in genes involved directly in insulin signaling, but rather those involved in adipose tissue development or function (34). It is thus plausible that genetic defects severely impairing mitosis or response to replication stress ultimately produce insulin resistance through progressive failure to maintain one or more insulin responsive tissues, a notion in keeping with the paucity of adipose tissue and muscle we observe in our patients. This idea remains to be validated; however, there is growing interest in the link between pandemic forms of insulin resistance, oxidative stress, and thereby DNA damage (28), and so in vivo modeling of severe NSMCE2/MMS21 hypomorphism may not only aid understanding of this rare and extreme condition, but may also offer wider, paradigmatic insights into common disease.

Methods

Exome sequencing. Genomic DNA was extracted from blood leukocytes and prepared as described previously (35) for capture. Amplification and sequencing were performed on the Illumina Genome Analyzer II (index case) or Hi-Seq platform (parents) (Illumina) as 54- or 75-bp paired-end reads respectively. BWA software package 0.5.9, Picard v1.46, SamTools v0.1.16 (36), Dindel v1.01 (37), Genome Analysis Toolkit (GATK) v1.1-20 (38), and Variant Effect Predictor (VEP) v2.1.39 were used to map sequence reads to the genome reference sequence (NCBI build 37), to calculate read quality, to call single nucleotide variants (SNVs) and insertion/deletions, and to annotate against Ensembl build 64. Variants were filtered to include only those with a minimum depth of 4× to a maximum depth of 1200× and were reported in variant calling file (VCF) format v4.1. For the index case, 6.4 Gb of sequence was produced at a mean depth of 60, with 80.4% coverage at a depth greater than or equal to 10, using high mapping quality alignments only (i.e., mapping quality ≥ 30). Equivalent figures for the father and mother were 8.8 Gb, 95% and 86.1%, and 8.9 Gb, 97% and 91.0% respectively. Raw exome sequence data from the trio are available from the European Genome-phenome Archive (EGAN00001101658, EGAN00001001829, EGAN00001001831).

De novo and compound heterozygous mutation analysis. Exome data for PI and her parents were merged, and variants with a high posterior probability of being de novo (cutoff value 0.8) were extracted using DeNovoGear v0.5 (https://sourceforge.net/projects/denovogear/) (40). Python scripts were used to extract potential compound heterozygous mutations. Variant annotation has previously been described in detail (41). Potentially causal variants identified were reconfirmed by Sanger sequencing of PCR products using the BigDye Terminator v3.1 Cycle Sequencing Kit according to the manufacturer’s instructions.
(Applied Biosystems; Life Technologies) and run on ABI 3730 DNA Sequencers (Applied Biosystems). Sequence analysis was performed using the Mutation Surveyor software package v2.3.9 (SoftGenetics). For P2, all exons and exon-intron boundaries of NSMCE2 were sequenced using Sanger sequencing as above (for primer sequences, see Supplemental Table 1).

**Haplotype analysis.** Twenty-four SNPs were selected from HapMap release 27 (42) to capture the variation within the 2 linkage blocks surrounding the p.Ser116Leufs*18 and p.Ala234Glufs*4 mutations (chr8: 125672673-126874287). These were genotyped along with the microsatellites D8S1179 and D8S266 in both probands, their families and 54 ethnically matched controls predominantly from the west of Great Britain from the 1000 genomes GBA panel (http://www.1000genomes.org/about#ProjectSamples). SNP genotyping was performed using the iPLEX Gold Assay (Sequenom Inc.) according to the manufacturer’s instructions, while microsatellites were amplified by PCR using FAM-labeled forward primers, diluted, mixed with CC5 Internal Lane Size Standard 500 (Promega), and run on an ABI 3730 DNA Analyser (Applied Biosystems). Allelic determination was undertaken using GeneMapper software (Applied Biosystems). Of the 24 SNPs, 3 were monomorphic in our data set.

Analysis was performed on the 21 polymorphic SNPs using SHAPEIT2 (http://www.shapeit.fr/) to phase the genotypes and calculate the most likely haplotypes, and the microsatellite markers were used to further define the haplotypes where ambiguous.

**Patient cell culture.** Epstein Barr virus–transformed LCLs were generated from primary lymphocytes using standard procedures and maintained in RPMI 1640 supplemented with L-glutamine, antibiotics (penicillin streptomycin), and 15% FBS. Bloom syndrome LCLs generated from primary lymphocytes using standard procedures and maintained in MEM supplemented with L-glutamine, antibiotics (penicillin streptomycin), and 15% FBS. Patient cell culture.

**Establishment of cell lines conditionally expressing WT or ligase inactive SMCE2/MMS21.** 10 µg of pSLIK WT or SUMO LD NSMCE2 plasmid, 75 µg of each of the packaging plasmids pMDLg/pRRE and pRSREV, and 5 µg of the pseudotyping pVSV-G plasmid were cotransfected into HEK 293T cells on 10-cm dishes using CalPhos Mammalian Transfection Kit (Clontech). The culture medium was replaced 12 hours after transfection with BioWhittaker Ultraculture medium (Lonza). Viral supernatants were harvested at 24, 48, and 72 hours after transfection and concentrated by centrifugation in Centricron Plus-70 Ultraceł PL-100 (Millipore). Lentiviral particles were added at low MOI to dermal fibroblasts in the presence of 8 µg/ml of polybrene for 6 hours followed by selection in G418 (600 µg/ml; Sigma-Aldrich) for 3 weeks. Stable cell lines were maintained in the presence of 200 µg/ml G418.

**Mircronucleus assay.** Dermal fibroblasts were grown to 70% to 80% confluence ± 1 mg/ml of DOX, trypsinized, resuspended, counted, and seeded onto glass coverslips. Twenty hours later, cells were exposed to 1 mM HU or saline for 4 hours, followed by saline washing, and allowed to recover in the presence of 3 µg/ml of cytochalasin-B for 48 hours, fixed with 4% paraformaldehyde for 15 minutes at room temperature, permeabilized in 0.05% Triton X-100, and stained with Actin-stain (Invitrogen) in PBS. Micronucleus studies in LCLs, cells were treated with 1 mM HU for 4 hours; HU was washed out and cells cultured in cytochalasin-B for 24 hours prior to harvesting to generate binucleates. Indirect immunofluorescence using anti-β-tubulin antibodies (Santa Cruz Biotechnology Inc.) and counterstaining with DAPI (Invitrogen) in PBS. Cells were examined at ×100 magnification using a wide-field fluorescence microscope (BX61; Olympus). More recently, scoring criteria were according to Fenech (20), using 1,000 binucleates per experimental condition.

For micronucleus studies in LCLs, cells were treated with 1 mM HU for up to 6 hours, and the proportion of cells in S phase was measured using BrdU pulse-mediated flow cytometry. Cells were pulse labeled with 20 µM BrdU 20 minutes prior to harvesting, then fixed in ice-cold 70% ethanol for 24 hours and resuspended in PBS containing 0.5% Tween-20, 10 µg/ml propidium iodide, and 500 µg/ml RNase A. FITC-conjugated anti-BrdU (clone 44, BD) was used for detection of S phase cells following HCl denaturation. To monitor the rate of S phase progression LCLs were pulse labeled with 20 µM BrdU, trypsinized, resuspended, counted, and seeded onto glass coverslips. Twenty hours later, cells were exposed to 1 mM HU or saline for 4 hours, followed by saline washing, and allowed to recover in the presence of 3 µg/ml of cytochalasin-B for 48 hours, fixed with 4% paraformaldehyde for 15 minutes at room temperature, permeabilized in 0.05% Triton X-100, and stained with Actin-stain (Invitrogen) in PBS. Cells were examined at ×100 magnification using a wide-field fluorescence microscope (BX61; Olympus). More recently, scoring criteria were according to Fenech (20), using 1,000 binucleates per experimental condition.

For micronucleus studies in LCLs, cells were treated with 1 mM HU for up to 6 hours, and the proportion of cells in S phase was measured using BrdU pulse-mediated flow cytometry. Cells were pulse labeled with 20 µM BrdU 20 minutes prior to harvesting, then fixed in ice-cold 70% ethanol for 24 hours and resuspended in PBS containing 0.5% Tween-20, 10 µg/ml propidium iodide, and 500 µg/ml RNase A. FITC-conjugated anti-BrdU (clone 44, BD) was used for detection of S phase cells following HCl denaturation. To monitor the rate of S phase progression LCLs were pulse labeled for 30 minutes with BrdU, washed in PBS, and then incubated in complete medium supplemented with Colcemid (100 µg/ml). Cells were harvested at 2-hour intervals and processed as described above. Data were collected using a BD FACS Canto and analyzed with FACS DIVA software.

**SUMOylation assay.** 4 x 10^5 AS49 cells were cotransfected with 4 µg GFP-SUMO-1 plasmid (pEGFP–SUMO-1) and 10 µg of WT or mutant pCi-Neo-Myc-NSMCE2. Cells were harvested 24 hours after transfection and solubilized with immunoprecipitation (IP) buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM EDTA, 2 mM EGTA, 25 mM NaF, 25 mM β-glycerophosphate, 0.1 mM sodium orthovanadate, 0.2% Triton X-100, 0.3% NP40, and Promega protease inhibitors cocktail) containing 20 mM N-ethylmaleimide. Myc-tagged proteins were immunoprecipitated using Red Anti-Myc Affinity Beads (E6654; Sigma-Aldrich). SUMOylation of immunoprecipitated material was detected by immunoblotting using anti-GFP antibody (2555; Cell Signaling).

**FACS analysis.** LCLs were treated with 250 µM HU for up to 6 hours, and the proportion of cells in S phase was measured using BrdU pulse-mediated flow cytometry. Cells were pulse labeled with 20 µM BrdU 20 minutes prior to harvesting, then fixed in ice-cold 70% ethanol for 24 hours and resuspended in PBS containing 0.5% Tween-20, 10 µg/ml propidium iodide, and 500 µg/ml RNase A. FITC-conjugated anti-BrdU (clone 44, BD) was used for detection of S phase cells following HCl denaturation. To monitor the rate of S phase progression LCLs were pulse labeled for 30 minutes with BrdU, washed in PBS, and then incubated in complete medium supplemented with Colcemid (100 µg/ml). Cells were harvested at 2-hour intervals and processed as described above. Data were collected using a BD FACS Canto and analyzed with FACS DIVA software.
BLM foci. LCLs were swollen in 75 mM KCl (10 minutes), immobilized on poly-L-lysine–coated slides by cytospinning (Cytospin; Shandon), fixed with 3% PFA and 2% sucrose in PBS, permeabilized using 0.1% Triton X-100 in 5% BSA-PBS for 2 minutes, blocked (5% BSA-PBS for 10 minutes) and incubated with goat polyclonal anti-BLM (clone C-18, sc7790; Santa Cruz Biotechnology Inc.) at 1:50 dilution in 5% BSA for 1 hour. Cells were counterstained with the appropriate FITC-conjugated secondary antibody and the nuclei stained with DAPI. Primary fibroblasts were fixed with methanol (~20°C for 20 minutes) and BLM foci detected using anti-BLM antibodies (2742; Cell Signaling). Slides were analyzed using the Zeiss Axiosplan platform and images captured using Simple PCI software.

UV-induced SCE. LCLs were UV-irradiated (2 J/m²) and differential staining of sister chromatids was achieved by growing LCL in the presence of 25 μM BrdU for 48 hours. After 4 hours of treatment in Colcemid, cells were swollen in 75 mM KCl for 10 minutes and fixed overnight at 4°C with Carnoy’s solution (methanol/acetic acid, 3:1). Chromosomes were spread by dropping cells from 40 cm high onto frozen slides, which were then left to dry for 3 days. Slides were then incubated for 20 minutes in Hoescht (10 g/ml) in the dark. After washing in SSC buffer (2 M NaCl, 0.3 M tri-sodium citrate, pH 7.0) slides were irradiated for 1 hour under a UV lamp (355 nm), followed by incubation in SSC buffer at 60°C for 1 hour. Slides were finally stained with modified Giemsa stain (G5S00; Sigma-Aldrich).

Zebrafish studies. WT TLF zebrafish (Danio rerio) were maintained and raised using standard protocols (43) in accordance with UK Home Office guidelines. Two splice-blocking MOs, nsmce2 sp1 and nsmce2 sp2, were designed to knock down nsmce2 expression (GeneTools, Philomath) (Supplemental Table 4 and Supplemental Figure 5). nsmce2 sp1 and nsmce2 sp2 were tested independently for effectiveness (data not shown) and then used in combination at a 1:1 ratio (data reported here). One- to two-cell embryos obtained from natural matings were microinjected with 1.8 nl of a 0.55 mM nsmce2 sp1 and 2 cocktail solution or the same concentration of control MO (Supplemental Table 3). Uninjected embryos were kept for comparison.

The efficiency of knockdown was assessed by PCR on cDNA derived from morphant or control embryos collected at day 2. mRNA was extracted using Trizol (Invitrogen). One microgram of mRNA from each group was transcribed using Superscript II First Strand Synthesis System (Invitrogen). Specific primers were then used to amplify a region across exons 1-3 for nsmce2 sp1 and across exons 2-4 for nsmce2 sp2 as well as control primers for Agpat2 to check equal loading. Primers are detailed in Supplemental Table 3.

Embryos were cultured in egg water at 28.5°C for 48 hours before scoring phenotypes. Live fish images were captured using a Leica M80 stereomicroscope, and body length was calculated by the Leica Application Suite program. Embryos were assigned to categories defined with reference to the SD of the length of a population of uninjected WT fish from the same mating. The experiment was carried out 6 times. Data shown are representative. For rescue experiments capped full-length human NSMCE2 mRNA (either WT, S116Lfs*18, A234Efs*4, or SUMO LD) was transcribed in vitro following the manufacturer’s manual (mMessage mMMachine; Invitrogen). RNA was purified using a Megaclear column (Ambion) and quantified spectrophotometrically and the size verified on an agarose gel. 100 picograms of mRNA were coinjected with nsmce2/mms21 MO at the 1- to 2-cell stage and the phenotype scored after 48 hours.

Statistics. All numerical data were analyzed using parametric statistical tests, namely Student’s t test or ANOVA with post hoc Tukey’s testing, as appropriate. All tests used and thresholds for significance are indicated in figure legends.

Study approval. This study was approved by the UK National Research Ethics Committee. Written informed consent was obtained from all participants or their parents.

Acknowledgments

I. Barroso, R.K. Semple, D.B. Savage, and S. O’Rahilly were supported by the Wellcome Trust (grants WT098051, WT098498, WT091551, and WT095515, respectively), the Medical Research Council Centre for Obesity and Related Disorders, and the United Kingdom National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre. M. O’Driscoll is supported by Cancer Research UK, the UK Medical Research Council and Leukaemia Lymphoma Research (United Kingdom). Funding for human studies was also provided by the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant no. 115372). Thanks to Alan Lehmann and Elaine Taylor and Oskar Fernandez-Capatillo for providing antibodies. We are grateful for access to exome sequence data from the CoLaus cohort, sequenced through partnership between the Wellcome Trust Sanger Institute, the CoLaus principal investigators, and GlaxoSmithKline. We also thank the NHLBI GO Exome Sequencing Project and its ongoing studies, which provided exome variant calls for comparison, namely, the Lung GO (HL-102923), the WHI (HL-102924), the Broad GO (HL-102925), the Seattle GO (HL-102926), and the Heart GO (HL-103010) Sequencing Projects. This study also makes use of data generated by the UK10K Consortium derived from samples from TwinsUK and ALSPAC. See Supplemental Appendix for a list of the investigators who contributed to generation of data. UK10K was funded by the Wellcome Trust under award WT091310. Our thanks also to Emma Gray, David Jones (Sample Management), Danielle Walker (Sequencing Pipeline), Carol Scott and Jillian Durham (Variation Informatics), and the staff of the Wellcome Trust Sanger Institute. The authors are grateful for the involvement and assistance of both patients and their parents in this study.

Address correspondence to: Robert K. Semple, University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Box 289, Addenbrooke’s Hospital, Cambridge CB2 0QQ, United Kingdom. Phone: 44.1223.769.035; E-mail: rks16@cam.ac.uk. Or to: Inês Barroso, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SA, United Kingdom. Phone: 44.0.1223.496.928; E-mail: ib1@sanger.ac.uk. Or to: Mark O’Driscoll, Human DNA Damage Response Disorders Genome, Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9RQ, United Kingdom. Phone: 44.0.1273.877.515; E-mail: m.o-driscoll@sussex.ac.uk. Or to: Jean-Claude Carel, Université Paris Diderot, Sorbonne Paris Cité & INSERM CIE-5, Président CMEL Hôpital Universitaire Robert-Debré, 48, boulevard Séurier, 75935 Paris Cedex 19, France. Phone: 33.1.40.03.41.05; E-mail: jean-claude.carel@inserm.fr.