*Human Molecular Genetics*, 2017, Vol. 26, No. 22 4481–4493

**doi: 10.1093/hmg/ddx336**
Advance Access Publication Date: 30 August 2017
Original Article

ORIGINAL ARTICLE

# SETD6 dominant negative mutation in familial colorectal cancer type X

Lorena Martín-Morales[1], Michal Feldman[2,3], Zlata Vershinin[2,3], Pilar Garre[1], Trinidad Caldés[1,*] and Dan Levy[2,3,*]

[1]Molecular Oncology Laboratory, Department of Medical Oncology, Hospital Clínico San Carlos, IdISSC, CIBERONC, 28040 Madrid, Spain, [2]The Shraga Segal Department of Microbiology, Immunology and Genetics and [3]National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Be'er-Sheva 84105, Israel

*To whom correspondence should be addressed. Email: trinidad.caldes@salud.madrid.org (T.C.); or Email: ledan@post.bgu.ac.il (D.L.)



## Abstract

Familiar colorectal cancer type X (FCCTX) comprises families that fulfill the Amsterdam criteria for hereditary non-polyposis colorectal cancer, but that lack the mismatch repair deficiency that defines the Lynch syndrome. Thus, the genetic cause that increases the predisposition to colorectal and other related cancers in families with FCCTX remains to be elucidated. Using whole-exome sequencing, we have identified a truncating mutation in the *SETD6* gene (c.791_792insA, p.Met264IlefsTer3) in all the affected members of a FCCTX family. SETD6 is a mono-methyltransferase previously shown to modulate the NF-$\kappa$B and Wnt signaling pathways, among other. In the present study, we characterized the truncated version of SETD6, providing evidence that this SETD6 mutation may play a role in the cancer inheritance in this family. Here we demonstrate that the truncated SETD6 lacks its enzymatic activity as a methyltransferase, while maintaining other properties such as its expression, localization and substrate-binding ability. In addition, we show that the mutant allele is expressed and that the resulting protein competes with the wild type for their substrates, pointing to a dominant negative nature. These findings suggest that the identified mutation impairs the normal function of SETD6, which may result in the deregulation of the different pathways in which it is involved, contributing to the increased susceptibility to cancer in this FCCTX family.

## Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cancer-related cause of death in the world (1,2). It is estimated that familial risk is involved in up to 30% of all CRC cases (3,4), although not more than 5–6% are caused by known germline mutations in cancer-predisposing genes (5). Hereditary non-polyposis colorectal cancer (HNPCC) is the most common form of inherited CRC. HNPCC is a familial syndrome characterized by an increased incidence of CRC and other extracolonic tumors (6) that has been defined by the Amsterdam I and II clinical criteria (7,8). Approximately half of HNPCC families are referred to as Lynch syndrome, since they are explained by germline inactivating mutations in the mismatch repair (MMR) genes—including MLH1, MSH2, MSH6 and PMS2—or by a large deletion in EPCAM (located upstream of MSH2) (9,10). As a consequence, Lynch syndrome tumors lack the corresponding MMR proteins and fail to repair DNA through the mismatch repair pathway. This, in turn, causes microsatellite instability (MSI) and leads to the accumulation of somatic mutations. It is worth noting that although HNPCC and Lynch Syndrome used to be employed as synonyms, nowadays HNPCC is defined by the clinical criteria (Amsterdam I or II), while Lynch Syndrome refers to the families with MMR defects.

**Received:** August 6, 2017. **Revised:** August 6, 2017. **Accepted:** August 15, 2017

The other half of HNPCC families are MMR-proficient and present microsatellite-stable (MSS) tumors. These cases have been grouped under the term familial colorectal cancer type X (FCCTX), and the genetic basis underlying their cancer predisposition remains unknown (11). Several studies have reported that the tumors from FCCTX patients have different molecular and clinical features than both Lynch syndrome and sporadic CRC (12,13), suggesting the presence of other forms of genomic instability. FCCTX's tumors also show distinct gene expression profiles and deregulation of different canonical pathways, although some similarities have been described between FCCTX and MSS sporadic tumors (12,14,15). These alterations could result in the deregulation of genes involved in chromosomal segregation, genomic instability, apoptosis, proliferation, growth inhibition and migration (14,16). Nonetheless, FCCTX is a heterogeneous group of families, and we are still far from fully understanding the different events that may be involved in their tumor progression. In the same way, our knowledge regarding the genetic alterations that contribute to FCCTX's heredity is fairly limited. Although previous studies had identified a few genes involved in the cancer susceptibility of these FCCTX families (17–19), it was not until the arrival of Next-Generation Sequencing (NGS) that a larger amount of new cancer-predisposing genes are being discovered (20). In view of the published results, it seems that FCCTX does not form a single entity, since multiple different genes are thought to be involved in their cancer heritability (21). However, together they still explain the inheritance in only a small portion of these FCCTX families. Thus, the identification of new high-risk genes that contribute to the increased cancer susceptibility of FCCTX families remains to be a challenge and a priority.

Covalent post-translational modification (PTM) of proteins is key for the regulation of many biological processes (22). Among these modifications, lysine methylation plays a vital role in the regulation of many cellular signaling pathways (23,24). A lysine residue in a given protein can accept up to three methyl groups, forming mono-, di- and tri-methylated derivatives. The most studied is the methylation of histones, which can either suppress or activate gene transcription depending on which lysine is methylated (25). Similar to histones, non-histone proteins can be also targeted for methylation with diverse biological outcomes, such as changes in gene expression, protein stability and subcellular localization (26–34). Lysine methylation is catalyzed by protein lysine methyltransferases (PKMTs) that transfer a methyl group from a donor S-adenosyl-L-methionine (SAM) (35–39). Even though there are more than 70 PKMTs present in the human proteome, little is known about their enzymatic activity beyond histones.

SETD6 (SET domain-containing protein 6) is a member of the lysine methyltransferase family known to monomethylate RelA—a subunit of the transcription factor NF-κB—on lysine 310 (RelAK310me1). This methyl mark is recognized by GLP (a histone methyltransferase), which in turn methylates histone H3 on lysine 9, promoting a repressed chromatin state and thereby silencing the transcription of NF-κB target genes (40). In contrast, it was recently reported that SETD6 promotes RelA's transcriptional activity in bladder cancer (41). Additionally, SETD6 has been proven to participate in the canonical Wnt signaling cascade by forming a complex with PAK4 (p21-activated kinase 4) and the transcription factor β-catenin at the chromatin, leading to the activation of β-catenin target gene transcription (42). In recent studies, SETD6 has also been linked to the regulation of the nuclear hormone receptor signaling (43), embryonic stem cell differentiation (44) and oxidative stress response (45,46).

In the present study, we sequenced the whole exome from a FCCTX family with the aim of identifying new causative cancer-predisposing genetic variants. Among the different mutations that were detected, a rare frameshift mutation in the *SETD6* gene (resulting in a premature termination of the enzyme) was identified and further characterized, both in cell-free systems and in colon cancer cell line models. This mutation causes the loss of the C-terminal portion of SETD6, leaving an intact catalytic SET domain. Here we demonstrate that, while the truncated protein shows the same subcellular localization and substrate-binding properties as the wild type, it completely loses its catalytic activity. We also show that mutant SETD6 is expressed in the tumor and competes against the wild type in the ability to methylate its substrate, suggesting a dominant negative effect. Taken together, our findings support a model by which a dominant negative mutation in SETD6 may contribute to FCCTX's pathology.

## Results

### Whole-exome sequencing results from FCCTX family cc598

The whole exome was sequenced from two affected members and one healthy relative of a FCCTX family (II: 1, II: 2 and II: 3, Fig. 1A). The corresponding family—known as family cc598—fulfilled the Amsterdam I criteria (7,8), with four colorectal cancers affecting two successive generations and the earliest age at diagnosis being 34 years old. Supplementary Material, Table S1 shows in detail the clinical data from the different members of this family.

After rigorous filtering of the variants detected aiming to select shared rare heterozygote damaging variants, a number of candidate genes were prioritized. The candidate variants selected were: *SETD6* c.791_792insA (p.Met264IlefsTer3), *CCDC60* c.1558 C > T (p.Arg20*), *L3MBTL1* c.1125 A > G (p.Lys375=), *CCDC62* c.79 C > T (p.Arg27Trp) and *WDR33* c.3266 G > C (p.Arg1089Pro). All the candidate variants (detailed in Supplementary Material, Table S2) were present in the two affected members (II: 1 and II: 2) and absent in the healthy sibling (II: 3). However, among these, the most relevant changes in the protein were observed in *SETD6*.

### Rare *SETD6* variant cosegregates with CRC within family cc598

The *SETD6* variant was a frameshift mutation known as c.791_792insA, p.Met264IlefsTer3, or just M264Ifs*3 (rs768456822, ENST00000310682), and it affected the three main protein-coding transcripts of this gene (ENST00000310682.6, ENST00000219315.8 and ENST00000394266.8). This mutation consisted of an insertion of an adenine between positions 791 and 792 of the cDNA and, like the other candidate variants, was carried by the two affected members sequenced (II: 1 and II: 2, with CRC diagnosed at ages 64 and 56, respectively), while not present in the healthy sibling (II: 3, 76 years old) (Fig. 1). In addition, the segregation study confirmed that the mutation was also present in II: 4 (CRC diagnosed at 34), while absent in two other healthy relatives (III: 1 and III: 2, 51 and 49 years old, respectively) (Fig. 1 and Supplementary Material, Fig. S2). The segregation for the other candidate variants is shown in Supplementary Material, Table S3. Loss of heterozygosity (LOH) analysis of *SETD6* c.791_792insA revealed that LOH did not occur in the tumor of II: 2 (Supplementary Material, Fig. S2B).

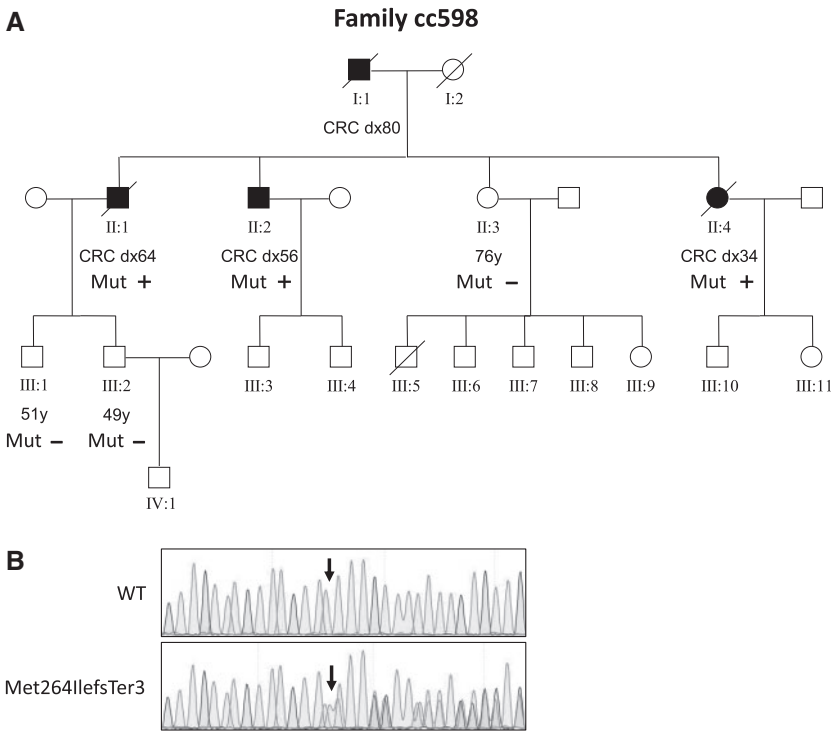


**Figure 1.** *SETD6* mutation c.791_792insA (p.M264Ifs*3) cosegregates with colorectal cancer in a FCCTX family. (**A**) Pedigree of family cc598 carrying the *SETD6* deleterious mutation c.791_792insA (p.Met264IlefsTer3). Whole-exome sequencing was done in family members II: 1, II: 2 and II: 3, and segregation was studied in members II: 4, III: 1 and III: 2. All affected participants were carriers (Mut+), while the healthy were non-carriers (Mut -). Colorectal cancer (CRC)-affected members are marked in black. The age at diagnosis (dx) or current age of healthy members is included beneath each individual (in years). (**B**) Electropherogram of the reverse wild-type and mutant sequence of the *SETD6* gene. The arrows show the point where the adenine is inserted.

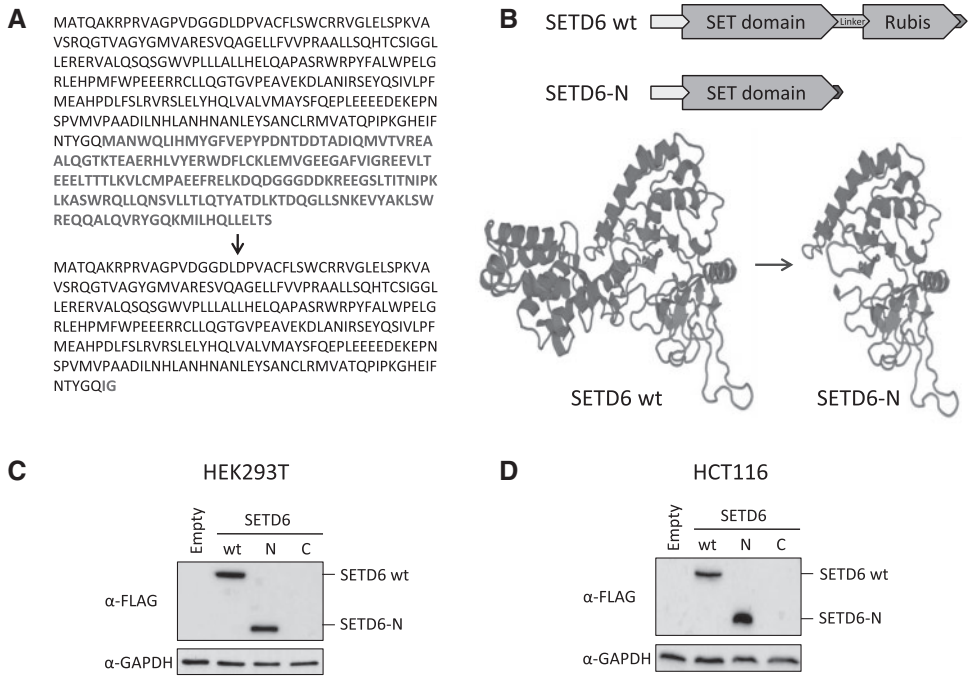


**Figure 2.** *SETD6* mutation causes the loss of the C-terminal half of the protein, but the mutant protein is still successfully overexpressed. (**A**) Amino acid sequence of SETD6 wt and SETD6-N with the changes marked in bold. (**B**) Representation of the different domains within the SETD6 protein showing the loss produced by the truncation (top), and the effect that the mutation is predicted to have on the 3D structure by SWISS-MODEL (53) (bottom). (**C, D**) Overexpression of FLAG-SETD6 (wt, N or C) in HEK293T (C) or HCT116 (D) cells. SETD6-N is the truncated version of the protein that mimics the mutation found in our family; SETD6-C is a truncated SETD6 that lacks the N-terminal part of the protein and the SET domain. The different versions of SETD6 were detected by western blot using an anti-FLAG antibody. GAPDH was used as a loading control.

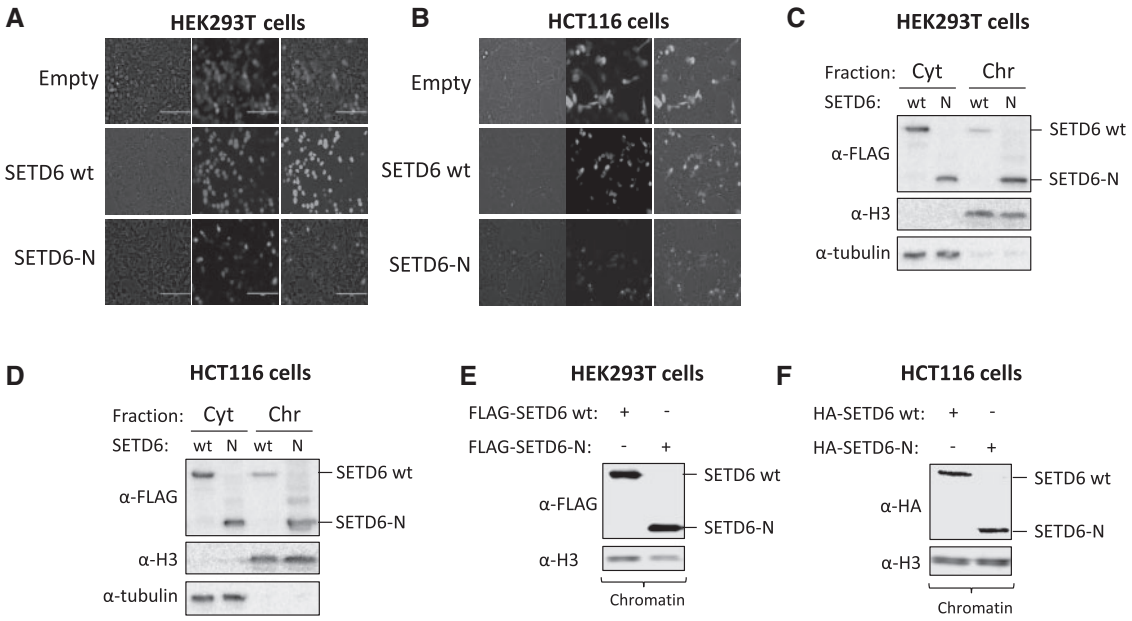


**Figure 3.** SETD6-N shows similar subcellular localization as wild-type SETD6. (**A, B**) HEK293T (A) and HCT116 (B) cells were transfected with GFP-tagged empty, SETD6 wt or SETD6-N plasmids. The distribution of GFP within the cells was observed by fluorescence microscopy. (**C, D**) HEK293T (C) or HCT116 (D) cells were transfected with FLAG-SETD6 (wt or N) and biochemically separated into cytoplasmic (Cyt) and chromatin (Chr) fractions. SETD6 was detected by western blot with an anti-FLAG antibody. Histone H3 and α-tubulin were used as chromatin and cytoplasmic controls, respectively. (**E, F**) The chromatin fraction of HEK293T (E) or HCT116 (D) cells transfected with FLAG-SETD6 (wt or N) was isolated and SETD6 was detected by western blot using the indicated antibodies. Histone H3 was used as a chromatin control.

On the other hand, although *SETD6* c.791_792insA (p.M264Ifs*3) is not a novel variant, it is quite rare in the general population, with a minor allele frequency (MAF) of 0.001285 according to ExAc and 0.00123 when the TCGA cohort is removed (47). Moreover, it is not found in the 1000 Genomes Project nor observed in homozygous state in neither of these databases. Finally, this variant produces a shift in the reading frame of the codons, and is therefore predicted to be Disease Causing by the MutationTaster software (48) (score = 1), further supporting its pathogenicity.

## SETD6 c.791_792insA (p.M264Ifs*3) results in the loss of the C-terminal half of the protein

As mentioned above, the identified *SETD6* mutation (Fig. 1B) results in a frameshift with the consequent introduction of two new amino acids at positions 264 and 265 (ENST00000310682)—Ile264 and Gly265, instead of the original Met and Ala—followed by a premature stop codon (Fig. 2A). The resulting truncated SETD6 has an intact SET domain but lacks the C-terminal half of the protein, which includes a linker sequence and the whole Rubisco domain, presumably responsible of mediating protein-protein interactions (49) (Fig. 2B). With the aim of checking whether mutant SETD6 is stable and normally expressed, a truncated version of the protein mimicking the frameshift mutation—hereon referred to as SETD6-N—was cloned and overexpressed. SETD6-N was found to be expressed to the same extent as wild-type SETD6 both in HEK293T and HCT116 cells (Fig. 2C and D, respectively). It is worth noting that an attempt to overexpress the C-terminal half of SETD6 (containing everything which SETD6-N is missing) was unsuccessful. Our working hypothesis was that the premature termination of the protein would impair SETD6's cellular function.

## SETD6-N shows the same localization pattern as wild-type SETD6

In order to test if the SETD6 mutation affects its subcellular localization, HEK293T and HCT116 cells were transfected with GFP-tagged SETD6 (either SETD6 wt or SETD6-N). Fluorescence microscopy of the transfected cells revealed that SETD6-N presented the same distribution within the cell as the wild type, both of which were mainly concentrated to the nucleus (Fig. 3A and B). We could detect, however, the presence of speckles inside the nucleus of the cells expressing mutant SETD6, which suggested that the association with the chromatin might be altered. To better understand the localization pattern, both proteins were transfected into cells followed by a biochemical fractionation of the cytosolic and chromatin fractions (50). This experiment confirmed a similar subcellular localization between the mutant and the wild-type, with no significant differences in the cytosolic fraction (Fig. 3C and D). Some differences between the wild-type and SETD6-N were only observed at the chromatin level. To further establish the presence of mutant SETD6 at chromatin, we used an additional method to extract the chromatin (see the Materials and Methods section for more details). This experiment confirmed comparable amounts of SETD6 wt and SETD6-N at this fraction (Fig. 3E and F). This data suggests that SETD6-N shows the same chromatin localization pattern as the wild-type enzyme.

## Recombinant SETD6-N binds its substrates to the same extent as wild-type SETD6

We next performed an ELISA experiment to test whether the studied mutation would affect the binding ability of SETD6 in a cell-free *in vitro* system. To this end, we compared the direct binding of the wild type and SETD6-N to two well-known

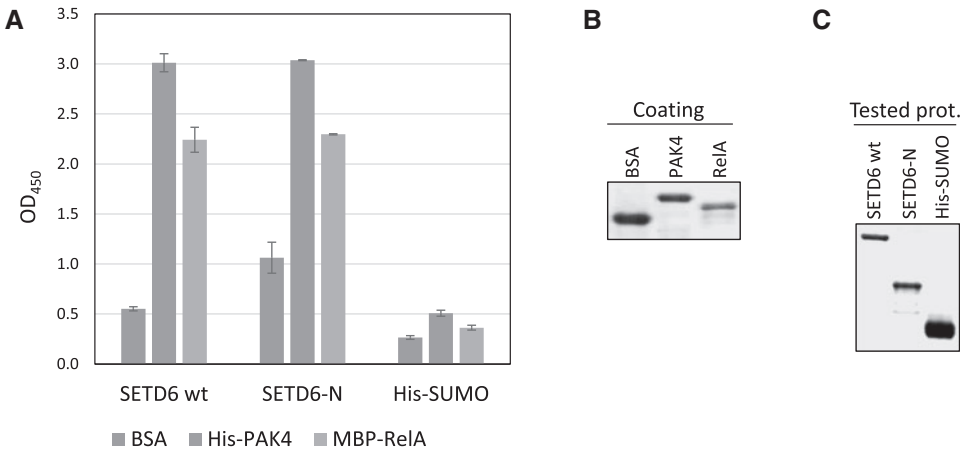


**Figure 4.** Recombinant SETD6-N shows similar binding to its substrates RelA and PAK4 as wild-type SETD6. (**A**) Interaction between recombinant SETD6 (wt or N) and PAK4/RelA determined by an ELISA. The plate was coated with 2 μg of His-PAK4, MBP-RelA or BSA (negative control), and then covered with 0.5 μg of His-SETD6 wt, His-SETD6-N or His-SUMO (negative control). Bound proteins were detected using a rabbit anti-SETD6 antibody. Data and error bars are from two technical replicates and represent two independent biological experiments. (**B, C**) Coomassie staining for the coating (B) and tested (C) proteins used in the ELISA.

substrates, RelA and PAK4 (40,42). Recombinant RelA and PAK4 were used to coat an ELISA plate, together with BSA, which was used as a negative control. Recombinant SETD6 wt, SETD6-N and SUMO were used as tested proteins, with the latter as another negative control. As shown in Figure 4A–C, SETD6-N bound at equal levels to both substrates as did wild-type SETD6, which suggests that the mutation does not impair the interaction between SETD6 and its substrates *in vitro*.

### Recombinant SETD6-N fails to methylate its substrates

To compare the enzymatic activity of wild-type SETD6 and SETD6-N, both proteins were subjected to a cell-free *in vitro* methylation assay with RelA and PAK4 as substrates (Fig. 5A and B, respectively). As expected, both substrates were specifically methylated by wild-type SETD6. However, no methylation was observed when SETD6-N was present in the reaction instead. It is of note that SETD6's automethylation activity was also lost in the truncated mutant. Consistent with these results, we found that while the recombinant wild-type SETD6 methylated immunoprecipitated RelA and PAK4 from HEK293T cells, SETD6-N failed to do so (Fig. 5C and D, respectively). In a reciprocal experiment, we observed that recombinant RelA and PAK4 were specifically methylated by wild-type SETD6 isolated from human cells and not by SETD6-N (Fig. 5E and F, respectively). These complementary assays further demonstrate that the SETD6 truncating mutation identified in the FCCTX family abrogates the enzymatic activity of SETD6.

### SETD6-N binds its substrates but loses its activity in colon cancer cells

Since the *SETD6* mutation was identified in hereditary colon cancer patients, we next aimed to confirm our findings in the colon cancer cell line HCT116. For this purpose, either wild-type SETD6 or SETD6-N were overexpressed in the presence or absence of FLAG-RelA (Fig. 6A). As expected, SETD6 wt and SETD6-N physically interacted with RelA at the chromatin to the same extent. Consistent with our cell-free *in vitro* experiments, we also observed that while SETD6 wt methylated RelA, SETD6-N did not. The methylation of RelA at the chromatin was identified using a RelAK310me1 antibody that could specifically recognize monomethylation of RelA at position K310 (40). The same results were obtained when PAK4 was used as the substrate (Fig. 6B). These findings support the dominant negative nature of *SETD6* c.791_792insA (p.M264Ifs*3) in a colon cancer cellular model.

### Both wild-type and mutant SETD6 are expressed in the tumor and compete for their substrates

Last but not least, we aimed to check if SETD6's mutant allele was expressed, given that the variant was carried in heterozygosis. To this end, a digital PCR was carried out using two TaqMan probes that specifically recognized either the wild-type or the mutant cDNA (with the insertion of the adenine). As observed in Figure 7A and B, the tumor from member II: 2 showed positive expression of both alleles. As expected, all non-carrier controls only expressed the wild-type allele.

The fact that both alleles were expressed raised the question of whether mutant and wild-type SETD6 would compete for their substrates. In order to address this issue, a cell-free *in vitro* methylation competition assay was performed in the presence of recombinant SETD6 wt and different amounts of recombinant SETD6-N, using RelA as the substrate. Figure 7C shows how wild-type SETD6's activity was inhibited in the presence of SETD6-N in a concentration-dependent manner, with a significant reduction in the amount of methylated substrate even when the same amount of each form of the enzyme was present (lanes 2 and 3, respectively). Finally, a competition assay in HCT116 cells confirmed the cell-free *in vitro* results, showing that the methylation of RelA at K310 in cells is reduced upon SETD6-N overexpression in a dose-dependent manner (Fig. 7D). Taken together, these results suggest that the expression of the mutant allele in the carriers is expected to compete with the enzymatic activity of SETD6 wt, supporting a dominant negative role of this mutation.

## Discussion

Familial colorectal cancer type X (FCCTX) is a term used to describe a heterogeneous group of CRC families for whom the genetic basis underlying their cancer predisposition remains unknown. Genome-wide analyses of gene expression patterns

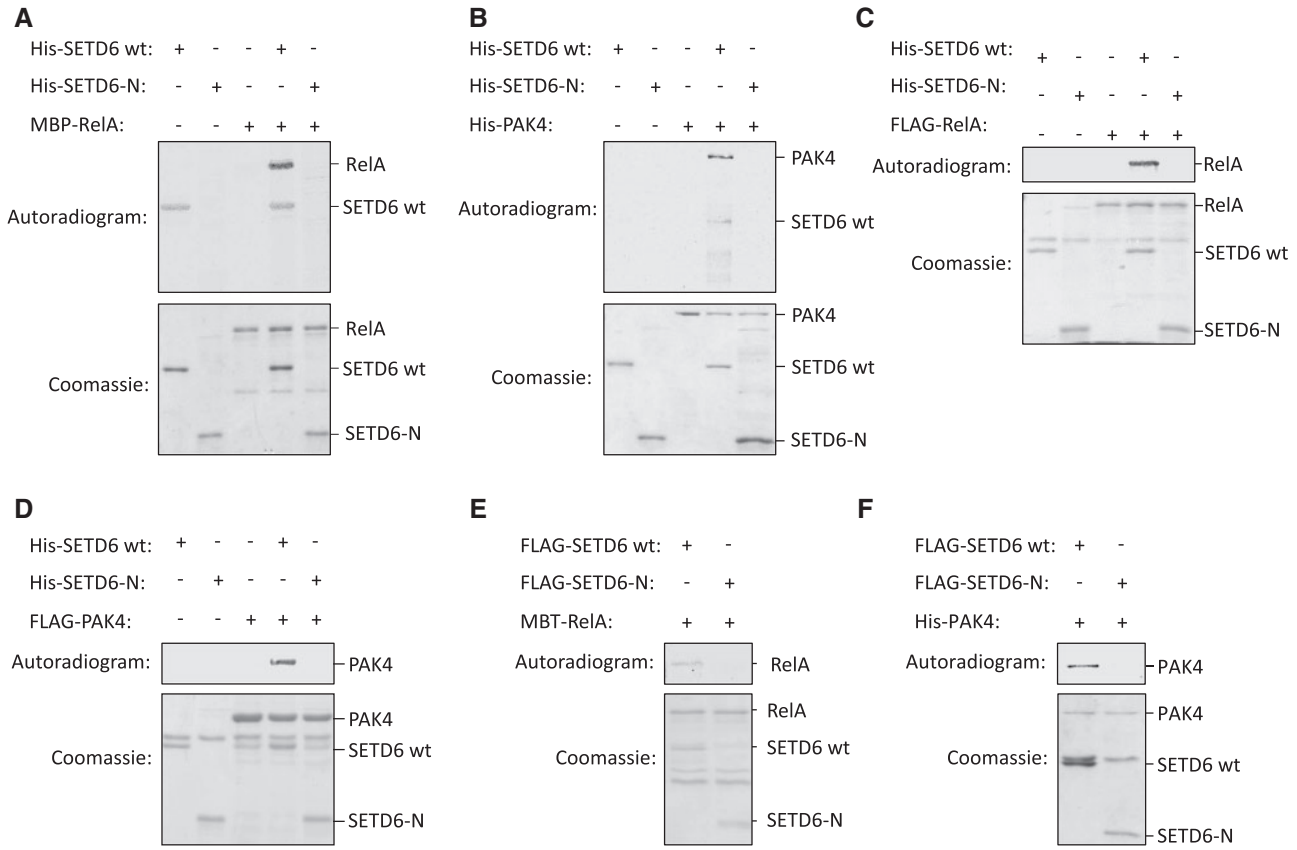


**Figure 5.** SETD6-N fails to methylate its substrates RelA and PAK4. (**A, B**) Cell-free *in vitro* methylation assays in the presence of $^{3}$H-SAM, recombinant His-SETD6 (wt or N) and either MBP-RelA (A) or His-PAK4 (B). The methylated proteins were detected by autoradiogram (top images), and the input used in the reactions is shown by Coomassie staining (bottom). (**C, D**) HEK293T cells were transfected with FLAG-RelA (C), FLAG-PAK4 (D) or empty plasmid. Cell extracts were immunoprecipitated with anti-FLAG M2 beads, followed by a radioactive cell-free *in vitro* methylation assay in the presence of $^{3}$H-SAM and recombinant His-SETD6 (wt or N). The methylated proteins were detected by autoradiogram (top images), and the input used in the reactions was shown by Coomassie staining (bottom). (**E, F**) HCT116 cells were transfected with either FLAG-SETD6 wt, FLAG-SETD6-N or empty plasmid. Cell extracts were immunoprecipitated with anti-FLAG M2 beads, followed by a radioactive cell-free *in vitro* methylation assay in the presence of $^{3}$H-SAM and recombinant MBP-RelA (E) or His-PAK4 (F). The methylated proteins were detected by autoradiogram (top images), and the input used in the reactions was shown by Coomassie staining (bottom).

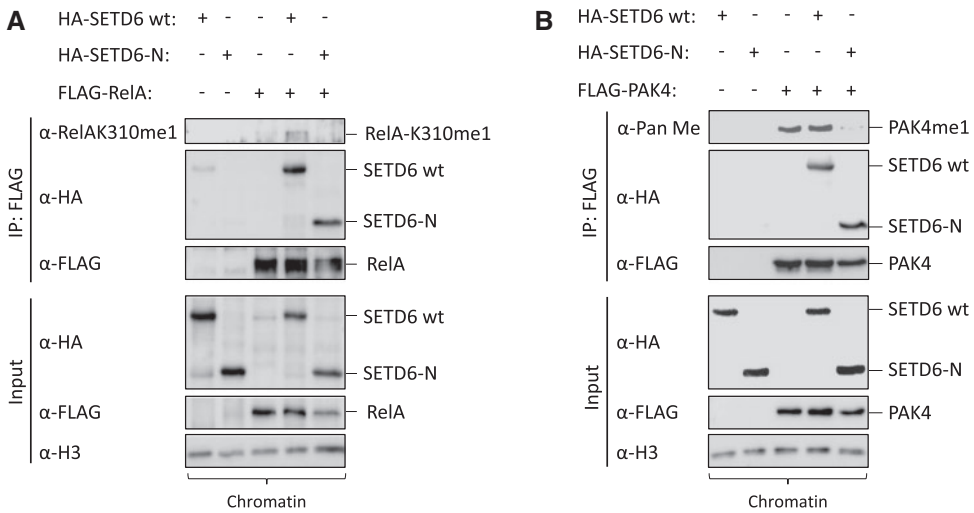


**Figure 6.** SETD6-N binds its substrates RelA and PAK4, while failing to methylate them, in the colon cancer cell line HCT116. (**A, B**) HCT116 cells were transfected with either HA-SETD6 wt or HA-SETD6-N plasmids, in the absence or presence of FLAG-RelA (A) or FLAG-PAK4 (B). The chromatin fraction was then immunoprecipitated with anti-FLAG M2 beads and analysed by western blot with the indicated antibodies.

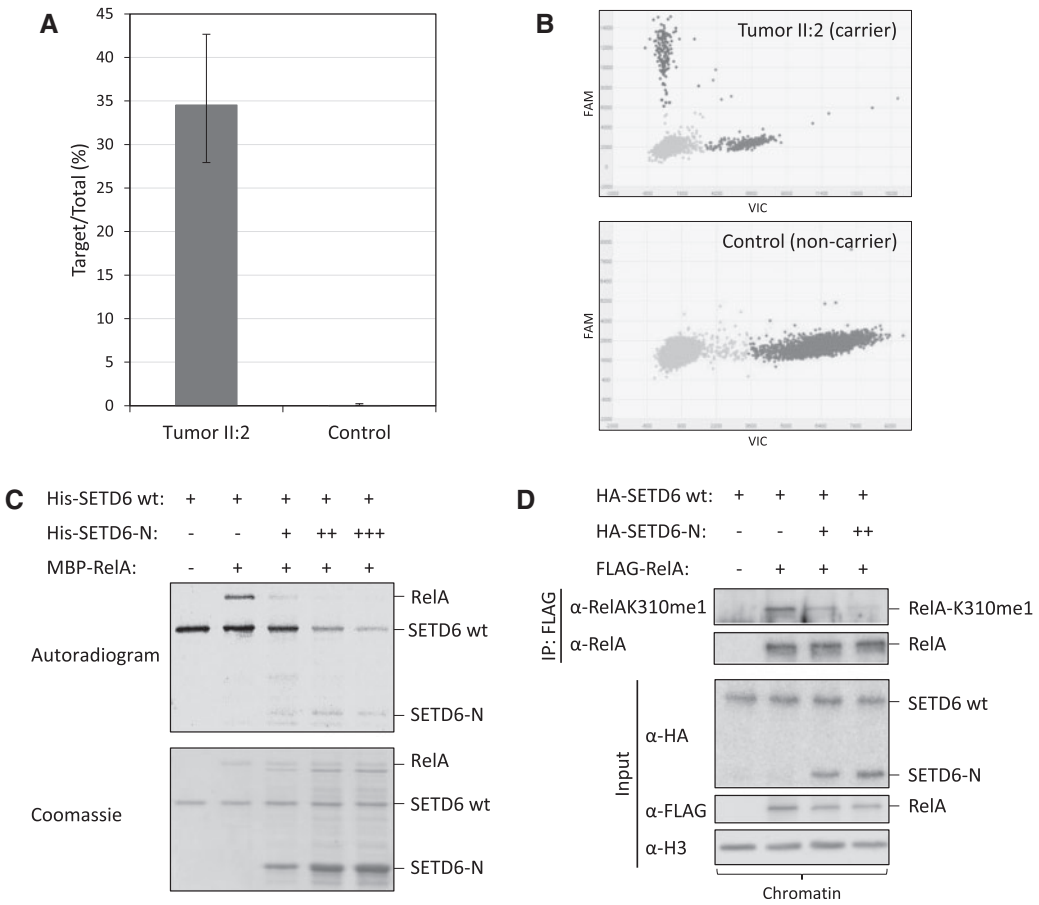


**Figure 7.** Both wild-type and mutant SETD6 are expressed in the tumor and compete for their substrates. (**A**) Allele-specific expression obtained by digital PCR presented as Target/Total, where "Target" is the mutant SETD6 allele (detected with the FAM dye). Data from the tumor of II:2 was collected from three independent experiments, and the error bars correspond to the confidence intervals. The Control represents tumor cDNA from four different sporadic CRC patients used as non-carrier controls. (**B**) Digital PCR visualization of the cDNA from the tumor of member II:2 (top) and a non-carrier control (bottom). The FAM dye detects the mutant allele, while the VIC dye detects the wild-type allele. (**C**) Radioactive cell-free *in vitro* methylation assay in the presence of recombinant wild-type SETD6 and different amounts (+ and ++) of recombinant SETD6-N. The methylated proteins were detected by autoradiogram (top), and the input used in the reactions is shown by Coomassie staining (bottom). (**D**) HCT116 cells were transfected with HA-SETD6 wt, FLAG-RelA and with or without different concentrations (+ and ++) of HA-SETD6-N. The chromatin fraction was then immunoprecipitated with anti-FLAG M2 beads and analyzed by western blot with the indicated antibodies.

in FCCTX in comparison to Lynch syndrome tumors (15) have shown differences in three major cancer-related pathways: cell cycle progression, oxidative phosphorylation and G protein-coupled receptor signaling, all of which have been previously linked to CRC (51–53). The fact that different genes are expressed in the tumors of FCCTX and Lynch syndrome patients may suggest that different molecular mechanisms mediate the progression of these two pathologies. FCCTX presumably comprises different yet-to-be-discovered genetic syndromes involving high-penetrance mutations in novel cancer-predisposing genes. However, it is thought that some FCCTX cases would result from a combination of low-penetrance mutations in different genes, or even from aggregation of sporadic cases due to shared lifestyle factors (5), what makes their study more difficult. The arrival of Next-Generation Sequencing (NGS) has been an important milestone in the search for new predisposition genes that explain the cancer heritability in these families. Nevertheless, this task is still a challenge and it is believed that the best strategy is to address each family individually, due to their wide diversity.

Using whole-exome sequencing, a rare frameshift mutation in the *SETD6* gene was identified in a FCCTX family. This mutation was found to cosegregate with the disease within the family, since it was carried by three CRC-affected members (II: 1, II: 2 and II: 4) while absent in three healthy relatives (II: 3, III: 1 and III: 2). No LOH was observed in the tumor of member II: 2, suggesting that in this family the *SETD6* mutation does not follow Knudson's two-hit mechanism for tumor suppressor genes, consistent with a dominant negative nature. Nonetheless, the LOH does not predict the pathogenicity of a mutation, since it has been reported to have a dual role in HNPCC families, with no preferential loss of the wild-type or the mutant allele. As a matter of fact, different carriers of well-known pathogenic MMR mutations have been reported to show the three different LOH statuses (no LOH, LOH of the mutant and LOH of the wild-type), even within the same family (54).

Regarding the consequences of the studied *SETD6* mutation, this variant results in the introduction of two amino acids at positions 264 and 265 followed by a premature stop codon. The truncated protein (SETD6-N) lacks its C-terminal half but retains an intact SET domain, which is responsible for its catalytic activity. Here, we show that mutant SETD6 exhibits dominant negative properties in cell-free *in vitro* systems and in a colon cancer *in vitro* cell line model. Although mutant SETD6 displays

similar localization, expression and binding to its known partners as the wild-type, the mutant protein loses its enzymatic activity. Indeed, unlike wild-type SETD6, SETD6-N lacks both its automethylation activity and the ability to methylate two previously identified substrates, PAK4 and RelA (40,42). In addition, the two alleles (wild type and mutant) were found to be expressed in the tumor of one of the carriers, and the two forms of SETD6 were shown to compete for their substrates both in cell-free systems and in a colon cancer cellular model, pointing to a dominant negative role. As a result, this mutation may have several downstream effects on the different pathways in which SETD6 is involved. For example, SETD6-dependent methylation of the NF-κB subunit RelA has been shown to be critical for basal inhibition of NF-κB signaling in the absence of stimulation (40). The NF-κB family of transcription factors has an essential role in inflammation and innate immunity, but it has also been increasingly recognized as a crucial player in many steps of cancer initiation and progression (55). In fact, activation of the NF-κB pathway has been positively associated with multiple types of cancer, including CRC (56). In addition, we had previously demonstrated that methylation of PAK4 by SETD6 promotes the activation of the Wnt/β-catenin pathway (42). On the other hand, deregulation of the Wnt signaling pathway has been shown to contribute to CRC development, including in HNPCC (57–60). Thus, the loss of SETD6's function, together with the aberrant regulation of NF-κB and/or Wnt signaling, could contribute to the initiation or progression of CRC in the studied FCCTX family. However, since SETD6 has several other substrates and is involved in numerous other signaling pathways (41,43,45,46), future studies are needed in order to define the downstream consequences of this SETD6 truncating mutation that we have identified.

Given all the results presented above, we propose that the presence of SETD6's mutant allele would presumably increase the cumulative risk of developing colorectal cancer throughout the life of the carriers, as compared with the general population. More research should be done in order to determine the penetrance of this mutation (that is to say the percentage of individuals who present this variant that will develop the disease, which is not complete even in well-known pathogenic MMR variants), whether it is a high, moderate or low-risk allele, and if there is also an association with other CRC-related cancers. Interestingly, the same SETD6 frameshift mutation had already been proposed to be associated with ovarian cancer in an extensive study by Kanchi and colleagues (61). Nonetheless, the search for new genes by exome sequencing in FCCTX families has demonstrated that more than one gene can be implicated in their increased risk of developing cancer. Actually, even when there is a high-risk gene involved, there might be also low-penetrance alleles that cooperate in the process as risk modifiers. Apart from these polygenic approaches, there are many other factors that might take part in modulating the cancer risk, including lifestyle and other environmental factors. For all these reasons, our results are not enough to claim that SETD6 alone is responsible for the increased CRC-risk in this FCCTX family. Hence, it is still difficult to know the effect that this mutation might have in the other carriers found in the public databases, as much as it is hard to predict if they are going to develop cancer at some point in their life or even if they have already developed it.

Noteworthy, the whole-exome sequencing followed by rigorous filtering identified additional candidate genes beside *SETD6* (Supplementary Material, Table S2). For instance, *CCDC60* showed a stop gain variant that results in the loss of 30 amino acids, 20 of which belong to a domain of unknown function, while the remaining 10 belong to a low complexity region. Another candidate was *L3MBTL1*, located on the long arm of chromosome 20 within a region that has been previously shown to be deleted in several malignancies (62). This gene presented a splicing variant that does not imply the gain or loss of a splicing acceptor/donor, but that is predicted to create a new exonic splicing silencer site instead. Finally, missense mutations were found in *CCDC62*, previously linked to prostate cancer (63), and in the polyadenylation regulating factor *WDR33* (64). Both variants affect low complexity regions of the corresponding proteins, and although they are predicted to be damaging *in silico*, they do not alter any known protein domains. While the current paper focuses on *SETD6*, we cannot exclude the possibility that the additional candidate variants identified may contribute independently or together to the pathology of FCCTX.

Together, our findings suggest that a truncating dominant negative mutation in SETD6 could potentially explain the cancer predisposition of this FCCTX family. These results certainly point to a pathogenic role of *SETD6* c.791_792insA (p.M264Ifs*3), though not enough to prove that *SETD6* alone is responsible for their increased cancer risk. Although no other *SETD6* variants were found in the remaining 12 families that were studied, nor in 109 familial CRC cases provided by Dr. Castellvi-Bel (Hospital Clinic, IDIBAPS, Barcelona, Spain), the screening of this gene in a larger group of patients could provide more insights into its role in other FCCTX families.

## Materials and Methods

### Study population

The studied family (cc598) was selected among a group of FCCTX families collected in the Genetic Counseling Unit at the Hospital Clínico San Carlos of Madrid. All FCCTX families fulfilled the Amsterdam I or II criteria for HNPCC (7) (at least three relatives with CRC or other related cancers, one being a first degree relative of the other two, with at least two successive generations involved, the earliest age of onset being 50 years old or lower and familial adenomatous polyposis excluded). In addition, all the CRC tumors from these families were MSS and presented normal expression of the MMR proteins. The study was approved by the Institutional Review Boards of the Hospital Clínico San Carlos, and an informed consent was obtained from each participant. Personal and family histories were obtained from the proband and participating relatives, and cancer diagnoses were confirmed by medical and pathology records.

### Family cc598

Family cc598 (Fig. 1A), selected for the whole-exome study, is an Amsterdam I family in which the father (I: 1) was diagnosed with CRC at the age of 80. He had two daughters and two sons, three of whom were diagnosed with CRC at ages 64 (II: 1), 56 (II: 2) and 34 (II: 4). Only one daughter was healthy (II: 3). The stage and location of the CRCs developed by II: 1, II: 2 and II: 4 were pT3N0M1 (splenic flexure, left colon), pT2N0M0 (right colon) and pT3N1M0 (rectum), respectively. The tumors from members II: 1 and II: 4 were MSS and showed normal expression of the MMR proteins (MLH1, MSH2, MSH6 and PMS2) (Supplementary Material, Table S1). The whole-exome sequencing was performed in germline DNA from members II: 1, II: 2 and II: 3.

### DNA and RNA extraction

Germline DNA and RNA were extracted from peripheral blood using the MagNA Pure Compact extractor system (Roche Diagnostics), according to the manufacturer's protocol. The PAXgene Blood RNA Kit (PreAnalytiX) was used to extract germline RNA conforming to its manual when the patient could not come to our hospital. Tumor DNA and RNA were extracted from 7 μm-thick haematoxylin/eosin-stained sections of the paraffin-embedded tumor tissues with a tumor content of more than 80% as determined by two experienced pathologists. Tumor DNA was extracted using the QIAamp DNA FFPE Tissue Kit from Qiagen, while tumor RNA was isolated employing the RNeasy FFPE Kit (Qiagen), according to their corresponding protocols. A NanoDrop® (ND1000) spectrophotometer was used to assess the DNA quantity and quality.

### Whole-exome sequencing

The whole-exome sequencing was outsourced to Sistemas Genómicos®. Library preparation for the capture of selected DNA regions was performed according to Agilent's SureSelect protocol for Illumina paired-end sequencing (SureSelectXT Human All Exon V3, 51 Mb, Agilent Technologies). The final library size and concentration were determined on an Agilent 2100 Bioanalyzer and a Qubit Fluorometer (Thermo Fisher Scientific), respectively. Finally, the library was sequenced on an Illumina HiSeq 2000 platform with paired-end reads of 101 bp, following the manufacturer's protocol. Images produced by the HiSeq 2000 were processed using the manufacturer's software to generate FASTQ sequence files. Reads were aligned against the human reference genome version GRCh37/hg19 using the BWA software, creating the BAM files. Low quality reads, PCR duplicates and other sequences that could introduce major biases were removed using Picard-tools (http://picard.sourceforge.net/) and SAMtools (65). Variant calling was performed using a combination of two different algorithms [VarScan (66) and GATK (67)] and the identified variants were annotated using the HGMD (68) and Ensembl (69) databases.

### Filtering and prioritization of the variants

The variants identified by whole-exome sequencing where subsequently filtered as follows: 1. Variants shared by the two affected members but not by the healthy relative were selected. 2. Homozygous variants were discarded, as well as variants present in allosomes. 3. Only coding non-synonymous (missense, stop gain, stop loss, in-frame, frameshift) and splicing variants were selected. 4. Variants with a minor allele frequency (MAF) in the general population greater than 0.01 were eliminated. 5. Missense and in-frame variants not predicted to be damaging by *in silico* programs [SIFT (70), Polyphen (71), MutationTaster (48)] were discarded, as well as splicing variants not directly affecting the donor/acceptor sites nor predicted to alter the splicing by the Human Splicing Finder (HSF) (72). Finally, the filtered variants were prioritized according to the genes and pathways involved.

### Variant validation, segregation and loss of heterozygosity (LOH) studies

All the candidate variants were validated by PCR followed by Sanger sequencing of the corresponding region of each gene, using specific primers that were designed with Primer3 (73). The segregation and LOH studies for the *SETD6* variant were also assessed by PCR and subsequent Sanger sequencing of the selected area of the *SETD6* gene (exon 7, ENST00000310682). The segregation study was carried out in germline DNA from the available members of the family (III: 1 and III: 2). However, although no germline DNA was available from the deceased member II: 4, we were able to study the segregation in this member extracting DNA from the paraffin-embedded tumor.

For the LOH analysis, tumor DNA was extracted from the paraffin-embedded tumor available, and the electropherograms of the germline and tumor sequences were compared, allowing the discrimination of the wild-type and the mutant alleles. LOH was considered when the intensity of any allele was reduced by $\geq$50% relative to the other allele. The *SETD6* primers used for the PCRs were CCACTCAGCCCATTCCTAAA (forward) and TGATACACTCACCCTGTAATGCT (reverse).

### Plasmids and cloning

The wild-type *SETD6* gene, as well as mutant *SETD6* (with the same variant identified in family cc598), was amplified by high-fidelity PCR and cloned into pcDNA3.1 and pET-Duet plasmids for overexpression and protein purification, respectively. The pcDNA3.1 plasmids in which the two forms of the gene were subcloned include pcDNA3.1-FLAG, pcDNA3.1-HA and pcDNA3.1-GFP, for the different experiments. In the same way, RelA and PAK4 were also cloned into pcDNA3.1-FLAG for the overexpression experiments (42) and pMAL-c2x or pET-Duet plasmids (respectively) for the expression and purification of the recombinant proteins.

### Cell lines and transfection

Two different cell lines were used in this study: human embryonic kidney cells (HEK293T) and human colon carcinoma cells (HCT116). Both were maintained in Dulbecco's modified Eagle's medium (Sigma-Aldrich, D5671) with 10% fetal bovine serum (Gibco), 2 mg/ml L-glutamine (Sigma-Aldrich, G7513), 1% penicillin-streptomycin (Sigma, P0781) and non-essential amino acids (Sigma-Aldrich, M7145), and they were cultured at 37 °C in a humidified incubator with 5% $CO_2$. For transient transfection, Mirus transfection reagents (*Trans*IT®-LT1 for HEK293T cells and *Trans*IT®-X2 for HCT116 cells) were used according to the manufacturer's instructions, together with Opti-MEM serum-free medium (Gibco).

### Western blot analysis

For western blot analyses, cells were homogenized and lysed in RIPA buffer [50 mM Tris-HCl, pH 8, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM DTT, and 1: 100 protease inhibitor mixture (Sigma-Aldrich)], except for the biochemical fractionation and chromatin immunoprecipitation experiments, in which the cells were lysed as described below. Samples were heated at 95 °C for 5 min in Laemmli sample buffer, run on a 8-12% SDS-PAGE electrophoresis gel, and then transferred to a polyvinylidene difluoride (PVDF) membrane. Membranes were blocked with either 10% skim milk in PBST or 5% BSA in TBST for 1 h on a shaking platform, and subsequently incubated with primary antibody for another hour with agitation. After three washes with the corresponding buffer (PBST or TBST), a 30-min incubation with HRP-conjugated secondary antibody and three additional washes, a 2-min reaction with a

chemiluminiscent substrate (EZ-ECL, Biological industries, 20-500-120) allowed the visualization of the proteins.

The mouse monoclonal antibodies used were: anti-FLAG M2 (Sigma-Aldrich, F1804), anti-HA (Millipore, 05–904), anti-GAPDH (Abcam, ab8245), anti-SETD6 (Genetex, GTX629891), and anti-histone H3 (Abcam, ab10799). The rabbit polyclonal antibodies used were: an HRP-conjugated pan methyl lysine antibody (ImmuneChem, ICP0502) and a specific antibody against mono-methylated RelA-Lys310 developed by Levy *et al.* (40). HRP-conjugated secondary antibodies (goat anti-rabbit and goat anti-mouse) were purchased from Jackson ImmunoResearch (111-035-144 and 115-035-062, respectively). Antibodies were diluted and prepared in PBST with 10% skim milk or in TBST with 5% BSA, according to the manufacturer's recommendations.

### Biochemical fractionation

Biochemical fractionation was performed as previously described by Mendez *et al.* (50), with the addition of a final resuspension of the chromatin pellet for 30 min on ice in RIPA buffer with 1 mM $MgCl_2$ and benzonase nuclease enzyme (Sigma-Aldrich). The chromatin fraction was obtained by the collection of the supernatant after low-speed centrifugation (5 min, 1700 g, 4 °C).

### Recombinant protein expression and purification

The *Escherichia coli* BL21-derivative Rosetta host strain was transformed with pET-Duet plasmids containing the gene of interest (*SETD6* wt, *SETD6-N*, *PAK4* or *RelA*) and grown overnight in 3 ml LB medium +100 μg/ml ampicillin (37 °C, 220 rpm). The culture was then expanded to 100 ml LB medium and incubated at 37 °C until the absorbance (OD) reached values of 0.6–0.8, when it was induced with 1: 10000 IPTG and left overnight at 18 °C and 220 rpm. After IPTG induction, the bacteria were harvested by centrifugation (10 min, 12000 rpm, 4 °C), resuspended in cold lysis buffer (10 mM imidazole, 1% PMSF, protease inhibitor cocktail, 0.1% triton and PBS) and then lysed by sonication on ice (25% amplitude, 1 min 30 s, 10 s on/5 s off). Finally, the lysate was centrifuged (20 min, 4 °C, 18000 rpm) and filtered, and the His-tagged proteins were purified using an ÄKTA™ column.

### ELISA

A high-binding 96-well polystyrene microplate (Greiner Bio-One MICROLON®) was coated with 2 μg of the recombinant proteins of interest (His-PAK4, MBP-RelA or BSA) diluted in PBS. The plate was blocked with 3% BSA in PBST and subsequently covered with 0.5 μg of the recombinant tested proteins (His-SETD6 wt, His-SETD6-N or His-SUMO as a control) diluted in 1% BSA in PBST. The rabbit polyclonal anti-SETD6 primary antibody (40) was then added, followed by incubation with an HRP-conjugated secondary antibody (goat anti-rabbit, 1: 2000; Jackson ImmunoResearch, 111-035-144). All the incubation steps were performed at room temperature with vigorous agitation for 1 h, and followed by three washes with PBST. After the final washes, 100 μl of TMB reagent were added to each well, succeeded after a few minutes by the same amount of 1 N $H_2SO_4$, in order to stop the reaction. The absorbance at 450 nm was then detected using an Infinite® M200 plate reader (Tecan). All samples were analysed in duplicates.

### Cell-free *in vitro* methylation assay

Cell-free *in vitro* methylation reactions with recombinant proteins took place in a final volume of 25 μl, containing 4 μg of substrate (His-PAK4 or MBP-RelA), 4 μg (or increasing amounts for the competition assay) of His-SETD6 (either wt or N), 2 mCi of $^3$H-labeled S-adenosyl-methionine (SAM) (AdoMet, Perkin-Elmer) and PKMT buffer (20 mM Tris-HCl, pH 8, 10% glycerol, 20 mM KCl, 5 mM $MgCl_2$). The reaction tubes were incubated overnight at 30 °C and then resolved by SDS-PAGE electrophoresis and subsequent autoradiogram. For the immunoprecipitation followed by cell-free *in vitro* methylation, human cells were transfected with empty, FLAG-SETD6 wt, FLAG-SETD6-N, FLAG-RelA or FLAG-PAK4 pcDNA3.1 plasmids, and 24 h post-transfection they were lysed with RIPA buffer and pulled down overnight with anti-FLAG M2 Affinity gel beads (Sigma-Aldrich, A2220) on a rotor at 4 °C. After two washes in RIPA buffer and another two in PKMT buffer, samples were subjected to an on-beads cell-free *in vitro* methylation assay as described above.

### Protein–protein chromatin immunoprecipitation

Protein-protein chromatin immunoprecipitation was modified from a published protocol (74). After cross-linking, cells were harvested and washed twice with PBS and then lysed in 1 ml of lysis buffer (20 mM Tris-HCl, pH 8, 85 mM KCl, 0.5% Nonidet P-40, and 1: 100 protease inhibitor mixture) for 10 min on ice. Nuclear pellets were resuspended in 200 μl of nuclei lysis buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS, 1: 100 protease inhibitor mixture) for 10 min on ice, and sonicated (Bioruptor, Diagenode) with high power settings for three cycles of 6 min each (30 s on/off). Samples were then centrifuged (20 min, 13000 rpm, 4 °C), and the soluble chromatin fraction was collected. The FLAG-labeled substrates present in the soluble chromatin were then immunoprecipitated overnight on a rotor at 4 °C, using 20 μl per tube of anti-FLAG M2 Affinity gel beads (Sigma-Aldrich, A2220). The beads were then washed according to the published protocol, heated for 30 min in Laemmli sample buffer at 95 °C, and resolved on 10-12% SDS-PAGE gels followed by western blot analyses.

### Reverse transcription polymerase chain reaction and digital PCR

Reverse transcription polymerase chain reaction (RT-PCR) was performed to convert RNA into complementary DNA (cDNA) using the PrimeScript RT Reagent Kit (Perfect Real Time) (Takara, Clontech), following the kit's instructions. For the expression assays, a TaqMan digital PCR (dPCR) was carried out taking advantage of the QuantStudio™ 3 D Digital PCR System (Thermo Fisher Scientific), according to the manufacturer's recommendations. The primers and TaqMan probes used in the dPCR were designed with the Custom TaqMan® Assay Design Tool (Thermo Fisher Scientific) and produced by the same company. The FAM probe specifically identified the mutant transcript, while the VIC probe only recognized the wild type. The dPCR was used to analyse the cDNA from the paraffin-embedded tumor of member II: 2. Tumor cDNA from four sporadic CRC patients were used as non-carrier controls.

## Supplementary Material

Supplementary Material is available at *HMG* online.

## Acknowledgements

We thank all the members from Levy's lab for their technical assistance, and Ruth Tennen and Ramon Birnbaum for critical reading of the manuscript. We also thank Sergi Calstellvi-Bel and Sebastia Franch-Exposito (Hospital Clinic, IDIBAPS, Barcelona, Spain) for the screening of *SETD6* in their familial CRC cohort.

*Conflict of Interest statement*. None declared.

## Funding

This work was supported by grants to DL from The Israel Science Foundation (285/14), The Research Career Development Award from the Israel Cancer Research Fund, Marie Curie Career Integration Grant and from the Israel Cancer Association. TC's and LMM's work was supported by grants PI-13/02588, PI-16/01292 and CIBERONC from the Carlos III Health Institute (Spain) and the European Regional Development FEDER funds. LMM's scientific stay in Ben-Gurion University was supported by the Federation of European Biochemical Societies (FEBS).

## References

1. Bray, F., Ferlay, J., Laversanne, M., Brewster, D.H., Gombe Mbalawa, C., Kohler, B., Pineros, M., Steliarova-Foucher, E., Swaminathan, R., Antoni, S. *et al*. (2015) Cancer incidence in five continents: inclusion criteria, highlights from Volume X and the global status of cancer registration. *Int. J. Cancer*, **137**, 2060–2071.
2. Steliarova-Foucher, E., O'Callaghan, M., Ferlay, J., Masuyer, E., Rosso, S., Forman, D., Bray, F. and Comber, H. (2015) The European Cancer Observatory: A new data resource. *Eur. J. Cancer*, **51**, 1131–1143.
3. Grady, W.M. (2003) Genetic testing for high-risk colon cancer patients. *Gastroenterology*, **124**, 1574–1594.
4. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl J. Med.*, **343**, 78–85.
5. Stoffel, E.M. and Kastrinos, F. (2014) Familial colorectal cancer, beyond Lynch syndrome. *Clin. Gastroenterol. Hepatol.*, **12**, 1059–1068.
6. Watson, P. and Lynch, H.T. (1993) Extracolonic cancer in hereditary nonpolyposis colorectal cancer. *Cancer*, **71**, 677–685.
7. Vasen, H.F., Mecklin, J.P., Khan, P.M. and Lynch, H.T. (1991) The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis. Colon Rectum.*, **34**, 424–425.
8. Vasen, H.F., Watson, P., Mecklin, J.P. and Lynch, H.T. (1999) New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology*, **116**, 1453–1456.
9. Lynch, H.T. and de la Chapelle, A. (1999) Genetic susceptibility to non-polyposis colorectal cancer. *J. Med. Genet.*, **36**, 801–818.
10. Lynch, H.T., Riegert-Johnson, D.L., Snyder, C., Lynch, J.F., Hagenkord, J., Boland, C.R., Rhees, J., Thibodeau, S.N., Boardman, L.A., Davies, J. *et al*. (2011) Lynch syndrome-associated extracolonic tumors are rare in two extended families with the same EPCAM deletion. *Am. J. Gastroenterol.*, **106**, 1829–1836.
11. Lindor, N.M., Rabe, K., Petersen, G.M., Haile, R., Casey, G., Baron, J., Gallinger, S., Bapat, B., Aronson, M., Hopper, J. *et al*. (2005) Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. *JAMA*, **293**, 1979–1985.
12. Sanchez-de-Abajo, A., de la Hoya, M., van Puijenbroek, M., Tosar, A., Lopez-Asenjo, J.A., Diaz-Rubio, E., Morreau, H. and Caldes, T. (2007) Molecular analysis of colorectal cancer tumors from patients with mismatch repair proficient hereditary nonpolyposis colorectal cancer suggests novel carcinogenic pathways. *Clin. Cancer Res.*, **13**, 5729–5735.
13. Garre, P., Martin, L., Bando, I., Tosar, A., Llovet, P., Sanz, J., Romero, A., de la Hoya, M., Diaz-Rubio, E. and Caldes, T. (2014) Cancer risk and overall survival in mismatch repair proficient hereditary non-polyposis colorectal cancer, Lynch syndrome and sporadic colorectal cancer. *Fam. Cancer*, **13**, 109–119.
14. Dominguez-Valentin, M., Therkildsen, C., Da Silva, S. and Nilbert, M. (2015) Familial colorectal cancer type X: genetic profiles and phenotypic features. *Mod. Pathol.*, **28**, 30–36.
15. Dominguez-Valentin, M., Therkildsen, C., Veerla, S., Jonsson, M., Bernstein, I., Borg, A. and Nilbert, M. (2013) Distinct gene expression signatures in lynch syndrome and familial colorectal cancer type x. *PLoS One*, **8**, e71755.
16. Therkildsen, C., Jonsson, G., Dominguez-Valentin, M., Nissen, A., Rambech, E., Halvarsson, B., Bernstein, I., Borg, K. and Nilbert, M. (2013) Gain of chromosomal region 20q and loss of 18 discriminates between Lynch syndrome and familial colorectal cancer. *Eur. J. Cancer*, **49**, 1226–1235.
17. Garre, P., Briceno, V., Xicola, R.M., Doyle, B.J., de la Hoya, M., Sanz, J., Llovet, P., Pescador, P., Puente, J., Diaz-Rubio, E. *et al*. (2011) Analysis of the oxidative damage repair genes NUDT1, OGG1, and MUTYH in patients from mismatch repair proficient HNPCC families (MSS-HNPCC). *Clin. Cancer Res.*, **17**, 1701–1712.
18. Garre, P., Martin, L., Sanz, J., Romero, A., Tosar, A., Bando, I., Llovet, P., Diaque, P., Garcia-Paredes, B., Diaz-Rubio, E. *et al*. (2015) BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. *Clin. Genet.*, **87**, 582–587.
19. Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I. *et al*. (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.*, **45**, 136–144.
20. Segui, N., Mina, L.B., Lazaro, C., Sanz-Pamplona, R., Pons, T., Navarro, M., Bellido, F., Lopez-Doriga, A., Valdes-Mas, R., Pineda, M. *et al*. (2015) Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology*, **149**, 563–566.
21. Esteban-Jurado, C., Vila-Casadesus, M., Garre, P., Lozano, J.J., Pristoupilova, A., Beltran, S., Munoz, J., Ocana, T., Balaguer, F., Lopez-Ceron, M. *et al*. (2015) Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet. Med.*, **17**, 131–142.
22. Pawson, T. and Warner, N. (2007) Oncogenic re-wiring of cellular signaling pathways. *Oncogene*, **26**, 1268–1275.
23. Greer, E.L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.*, **13**, 343–357.
24. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
25. Peterson, C.L. and Laniel, M.A. (2004) Histones and histone modifications. *Curr. Biol.*, **14**, R546–R551.

26. Carr, S.M., Munro, S., Zalmas, L.P., Fedorov, O., Johansson, C., Krojer, T., Sagum, C.A., Bedford, M.T., Oppermann, U. and La Thangue, N.B. (2014) Lysine methylation-dependent binding of 53BP1 to the pRb tumor suppressor. *Proc. Natl Acad. Sci. U S A*, **111**, 11341–11346.
27. Hamamoto, R., Saloura, V. and Nakamura, Y. (2015) Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat. Rev. Cancer*, **15**, 110–124.
28. Hamamoto, R., Toyokawa, G., Nakakido, M., Ueda, K. and Nakamura, Y. (2014) SMYD2-dependent HSP90 methylation promotes cancer cell proliferation by regulating the chaperone complex formation. *Cancer Lett.*, **351**, 126–133.
29. Carr, S.M., Munro, S. and La Thangue, N.B. (2012) Lysine methylation and the regulation of p53. *Essays Biochem.*, **52**, 79–92.
30. Chatterjee, S., Senapati, P. and Kundu, T.K. (2012) Post-translational modifications of lysine in DNA-damage repair. *Essays Biochem.*, **52**, 93–111.
31. Huang, J. and Berger, S.L. (2008) The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.*, **18**, 152–158.
32. Shi, X., Kachirskaia, I., Yamaguchi, H., West, L.E., Wen, H., Wang, E.W., Dutta, S., Appella, E. and Gozani, O. (2007) Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol. Cell*, **27**, 636–646.
33. Zhang, K. and Dent, S.Y. (2005) Histone modifying enzymes and cancer: going beyond histones. *J. Cell Biochem.*, **96**, 1137–1148.
34. Zhang, X., Wen, H. and Shi, X. (2012) Lysine methylation: beyond histones. *Acta Biochim. Biophys. Sin (Shanghai)*, **44**, 14–27.
35. Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A. *et al.* (2011) NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell*, **44**, 609–620.
36. Medjkane, S., Cock-Rada, A. and Weitzman, J.B. (2012) Role of the SMYD3 histone methyltransferase in tumorigenesis: local or global effects? *Cell Cycle*, **11**, 1865.
37. Richly, H., Lange, M., Simboeck, E. and Di Croce, L. (2010) Setting and resetting of epigenetic marks in malignant transformation and development. *Bioessays*, **32**, 669–679.
38. Shi, Y. and Whetstine, J.R. (2007) Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell*, **25**, 1–14.
39. Yeates, T.O. (2002) Structures of SET domain proteins: protein lysine methyltransferases make their mark. *Cell*, **111**, 5–7.
40. Levy, D., Kuo, A.J., Chang, Y., Schaefer, U., Kitson, C., Cheung, P., Espejo, A., Zee, B.M., Liu, C.L., Tangsombatvisit, S. *et al.* (2011) Lysine methylation of the NF-kappaB subunit RelA by SETD6 couples activity of the histone methyltransferase GLP at chromatin to tonic repression of NF-kappaB signaling. *Nat. Immunol.*, **12**, 29–36.
41. Mukherjee, N., Cardenas, E., Bedolla, R. and Ghosh, R. (2017) SETD6 regulates NF-kappaB signaling in urothelial cell survival: Implications for bladder cancer. *Oncotarget*, **8**, 15114–15125.
42. Vershinin, Z., Feldman, M., Chen, A. and Levy, D. (2016) PAK4 methylation by SETD6 promotes the activation of the Wnt/beta-catenin pathway. *J. Biol. Chem.*, **291**, 6786–6795.
43. O'Neill, D.J., Williamson, S.C., Alkharaif, D., Monteiro, I.C., Goudreault, M., Gaughan, L., Robson, C.N., Gingras, A.C. and Binda, O. (2014) SETD6 controls the expression of estrogen-responsive genes and proliferation of breast carcinoma cells. *Epigenetics*, **9**, 942–950.
44. Binda, O., Sevilla, A., LeRoy, G., Lemischka, I.R., Garcia, B.A. and Richard, S. (2013) SETD6 monomethylates H2AZ on lysine 7 and is required for the maintenance of embryonic stem cell self-renewal. *Epigenetics*, **8**, 177–183.
45. Chen, A., Feldman, M., Vershinin, Z. and Levy, D. (2016) SETD6 is a negative regulator of oxidative stress response. *Biochim. Biophys. Acta*, **1859**, 420–427.
46. Cohn, O., Chen, A., Feldman, M. and Levy, D. (2016) Proteomic analysis of SETD6 interacting proteins. *Data in Brief*, **6**, 799–802.
47. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60, 706 humans. *Nature*, **536**, 285–291.
48. Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods*, **11**, 361–362.
49. Trievel, R.C., Flynn, E.M., Houtz, R.L. and Hurley, J.H. (2003) Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nat. Struct. Biol.*, **10**, 545–552.
50. Mendez, J. and Stillman, B. (2000) Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell Biol.*, **20**, 8602–8612.
51. Bertucci, F., Salas, S., Eysteries, S., Nasser, V., Finetti, P., Ginestier, C., Charafe-Jauffret, E., Loriod, B., Bachelart, L., Montfort, J. *et al.* (2004) Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. *Oncogene*, **23**, 1377–1391.
52. Wilson, C.H., McIntyre, R.E., Arends, M.J. and Adams, D.J. (2010) The activating mutation R201C in GNAS promotes intestinal tumourigenesis in Apc(Min/+) mice through activation of Wnt and ERK1/2 MAPK pathways. *Oncogene*, **29**, 4567–4575.
53. Dunican, D.S., McWilliam, P., Tighe, O., Parle-McDermott, A. and Croke, D.T. (2002) Gene expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. *Oncogene*, **21**, 3253–3257.
54. Sanchez de Abajo, A., de la Hoya, M., van Puijenbroek, M., Godino, J., Diaz-Rubio, E., Morreau, H. and Caldes, T. (2006) Dual role of LOH at MMR loci in hereditary non-polyposis colorectal cancer? *Oncogene*, **25**, 2124–2130.
55. Hoesel, B. and Schmid, J.A. (2013) The complexity of NF-kappaB signaling in inflammation and cancer. *Mol. Cancer*, **12**, 86.
56. Moorchung, N., Kunwar, S. and Ahmed, K.W. (2014) An evaluation of nuclear factor kappa B expression in colorectal carcinoma: an analysis of 50 cases. *J. Cancer Res. Ther.*, **10**, 631–635.
57. Huang, J., Kuismanen, S.A., Liu, T., Chadwick, R.B., Johnson, C.K., Stevens, M.W., Richards, S.K., Meek, J.E., Gao, X., Wright, F.A. *et al.* (2001) MSH6 and MSH3 are rarely involved in genetic predisposition to nonpolypotic colon cancer. *Cancer Res.*, **61**, 1619–1623.
58. Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. and Kinzler, K.W. (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*, **275**, 1787–1790.
59. Rowan, A.J., Lamlum, H., Ilyas, M., Wheeler, J., Straub, J., Papadopoulou, A., Bicknell, D., Bodmer, W.F. and Tomlinson, I.P. (2000) APC mutations in sporadic colorectal tumors: A mutational "hotspot" and interdependence of the "two hits". *Proc. Natl Acad. Sci. U S A*, **97**, 3352–3357.

60. Abdel-Rahman, W.M., Ollikainen, M., Kariola, R., Jarvinen, H.J., Mecklin, J.P., Nystrom-Lahti, M., Knuutila, S. and Peltomaki, P. (2005) Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. *Oncogene*, **24**, 1542–1551.
61. Kanchi, K.L., Johnson, K.J., Lu, C., McLellan, M.D., Leiserson, M.D., Wendl, M.C., Zhang, Q., Koboldt, D.C., Xie, M., Kandoth, C. *et al*. (2014) Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.*, **5**, 3156.
62. Gurvich, N., Perna, F., Farina, A., Voza, F., Menendez, S., Hurwitz, J. and Nimer, S.D. (2010) L3MBTL1 polycomb protein, a candidate tumor suppressor in del(20q12) myeloid disorders, is essential for genome stability. *Proc. Natl Acad. Sci. U S A*, **107**, 22552–22557.
63. Chen, M., Ni, J., Chang, H.C., Lin, C.Y., Muyan, M. and Yeh, S. (2009) CCDC62/ERAP75 functions as a coactivator to enhance estrogen receptor beta-mediated transactivation and target gene expression in prostate cancer cells. *Carcinogenesis*, **30**, 841–850.
64. Schonemann, L., Kuhn, U., Martin, G., Schafer, P., Gruber, A.R., Keller, W., Zavolan, M. and Wahle, E. (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.*, **28**, 2381–2393.
65. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
66. Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K. and Ding, L. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*, **25**, 2283–2285.
67. 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.
68. Stenson, P.D., Mort, M., Ball, E.V., Howells, K., Phillips, A.D., Thomas, N.S. and Cooper, D.N. (2009) The Human Gene Mutation Database: 2008 update. *Genome Med.*, **1**, 13.
69. Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L. *et al*. (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
70. Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
71. 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.
72. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.*, **37**, e67.
73. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
74. Nelson, J.D., Denisenko, O. and Bomsztyk, K. (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.*, **1**, 179–185.
75. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. *et al*. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.
76. den Dunnen, J.T. (2016) Sequence Variant Descriptions: HGVS Nomenclature and Mutalyzer. *Curr Protoc Hum Genet*, **90**, 7 13 11–17 13 19.
77. Smigielski, E.M., Sirotkin, K., Ward, M. and Sherry, S.T. (2000) dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res*, **28**, 352–355.
78. Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*, **14**, 178–192.