# Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist *GREM1*

Emma Jaeger[1,6], Simon Leedham[1,2,6], Annabelle Lewis[1], Stefania Segditsas[1], Martin Becker[1], Pedro Rodenas Cuadrado[1], Hayley Davis[1], Kulvinder Kaur[3], Karl Heinimann[1,4], Kimberley Howarth[1], James East[2], Jenny Taylor[3], Huw Thomas[5] & Ian Tomlinson[1,3]

**Hereditary mixed polyposis syndrome (HMPS) is characterized by apparent autosomal dominant inheritance of multiple types of colorectal polyp, with colorectal carcinoma occurring in a high proportion of affected individuals. Here, we use genetic mapping, copy-number analysis, exclusion of mutations by high-throughput sequencing, gene expression analysis and functional assays to show that HMPS is caused by a duplication spanning the 3′ end of the *SCG5* gene and a region upstream of the *GREM1* locus. This unusual mutation is associated with increased allele-specific *GREM1* expression. Whereas *GREM1* is expressed in intestinal subepithelial myofibroblasts in controls, *GREM1* is predominantly expressed in the epithelium of the large bowel in individuals with HMPS. The HMPS duplication contains predicted enhancer elements; some of these interact with the *GREM1* promoter and can drive gene expression *in vitro*. Increased *GREM1* expression is predicted to cause reduced bone morphogenetic protein (BMP) pathway activity, a mechanism that also underlies tumorigenesis in juvenile polyposis of the large bowel.**

Colorectal cancers that arise as a result of high-penetrance germline mutations in genes such as *APC*, *MSH2* and *MLH1* are associated with tumors of particular morphology and with variable but specific extra-colonic features. HMPS is an unusual disease, in that affected individuals can develop polyps of multiple and mixed morphologies—including serrated lesions, Peutz-Jeghers polyps, juvenile polyps, conventional adenomas and colorectal carcinoma (CRC)[1]—in the absence of any identifiable extra-colonic features. A single large Jewish family with HMPS (SM96) with dominantly inherited polyposis was initially identified at St Mark's Hospital, London, and germline mutations in known CRC predisposition genes were excluded over time[2]. Using linkage analysis in SM96 and independently in another Jewish kindred with polyposis, SM1311, we mapped the causative gene to *CRAC1*

(also known as *HMPS*) on chromosome 15q13.3 and showed that the two families shared a haplotype within this region[2,3]. Subsequently, we identified additional kindreds with HMPS on the basis of their clinical features and Ashkenazi descent and showed that all families shared a disease haplotype between 30.735 Mb and 31.370 Mb on chromosome 15q13.3 (ref. 4). To date, no non-Ashkenazi families with HMPS have convincingly been identified, although there is overlap between some of the clinical features of HMPS and those of juvenile polyposis syndrome (JPS)[5] and hyperplastic polyposis syndrome (HPPS)[6].

The possibility that a large proportion of HMPS cases was caused by a single founder mutation led us to hypothesize that the underlying genetic change might be somewhat unusual: we reasoned that disease-causing mutations might be of a very specific type that occurred infrequently as spontaneous events. Consistent with this notion, we found no convincing evidence of loss of heterozygosity at 15q13.3 in 12 tumors from individuals with HMPS (data not shown), suggesting that the gene was not a classical tumor suppressor. We also sequenced three subjects with HMPS and two unaffected relatives who did not share the disease haplotype for the coding regions of the three genes (*GREM1*, *SCG5* and *FMN1*) that lie within the region of the shared haplotype. We found known polymorphisms but no novel, potentially pathogenic changes that were consistently present in affected subjects and absent in their unaffected relatives.

We therefore wondered whether the underlying mutation in HMPS might take the form of a copy-number change. We designed a custom oligonucleotide array to search for copy-number variation in the region. This analysis showed the presence of a heterozygous single-copy duplication of approximately 40 kb centered on chr. 15: 30.77 Mb in two subjects with HMPS and in none of their three unaffected relatives (**Fig. 1** and **Supplementary Fig. 1**). PCR amplification across the duplication breakpoints subsequently mapped it to chr. 15: 30,752,231–30,792,051. The change was found to be a simple tandem

Figure 1 The HMPS duplication. (**a**) Array comparative genomic hybridization (CGH) analysis of the region around the HMPS duplication. The region of copy-number gain, estimated to be at a 3:2 ratio relative to controls and, hence, resulting from a single-copy gain, is indicated. Results from the full region analyzed are shown (**Supplementary Fig. 1**). (**b**) Schematic of the 40-kb duplication showing the involvement of the 3′ half of *SCG5* extending to a region upstream of *GREM1*. The locations on chromosome 15 of coding sequences, introns, selected SNPs and CpG islands are shown for wild-type and mutated copies of the chromosome. Arrows show the direction of transcription.
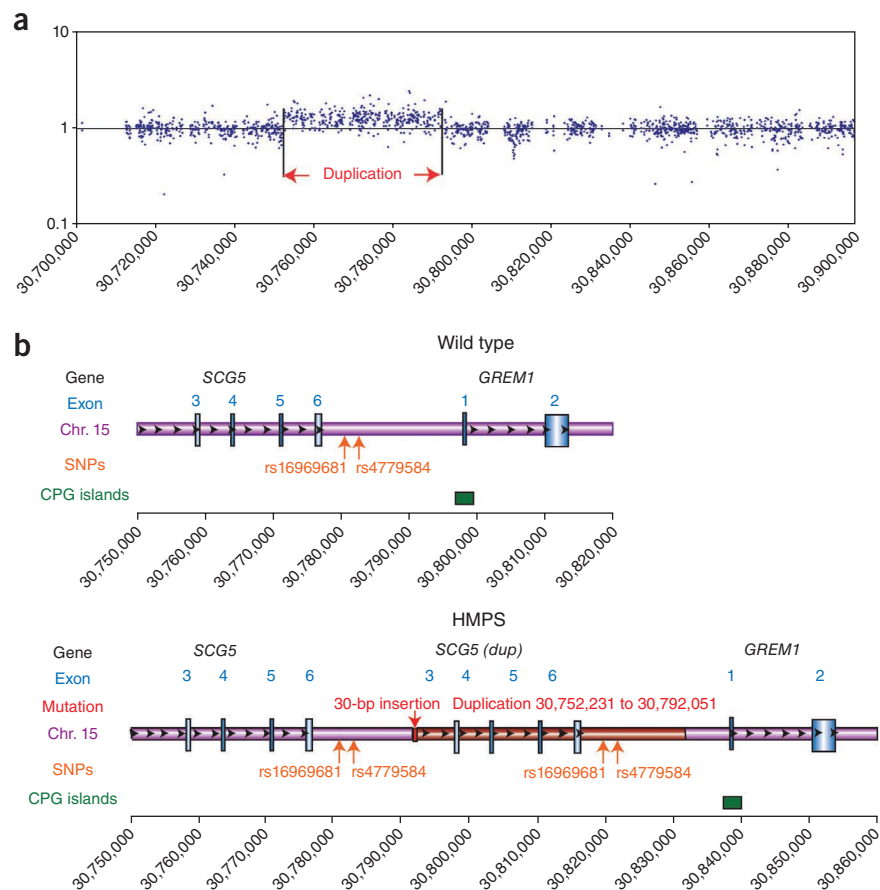
tail-head duplication with the insertion of a 30-bp sequence of unknown origin and no homology to known sequences between the duplicons. The duplication extended from intron 2 of *SCG5* to a site just upstream of the *GREM1* CpG island (**Fig. 1**). In order to investigate the duplication further, we designed PCR primers that spanned the duplication boundary and produced a unique, specific amplification product of 190 bp. As a control, we chose a region upstream of *GREM1* that was present in all individuals, which generated a product of 435 bp. We tested 40 affected individuals (**Supplementary Fig. 2**) and 50 unaffected individuals (either polyp-free but at-risk persons who were >40 years old or the spouses of affected individuals) from 6 putative families with HMPS. There was perfect concordance between presence of the 190-bp product and affected status (**Supplementary Fig. 1** and **Supplementary Table 1**). We then tested 188 unselected Ashkenazi controls, and none showed amplification of the 190-bp product; this group included one duplication-negative individual who shared the HMPS haplotype, suggesting that the duplication had arisen on that haplotype background.

We then wondered whether duplication of chromosome 15q13.3—either identical to that in HMPS or a similar change—might be present in a set of familial CRC cases from the general population of the UK. We therefore used the QuantiSNP program[7] to analyze our existing Illumina Hap550 SNP array data from 718 CRC cases and 935 controls enrolled in the Colorectal Tumour Gene Identification (CORGI) study of CRC genetics[8]. One case, but no control, was found to have a duplication close to *GREM1* and *SCG5*, and we showed that this change was identical to that found in the HMPS cases. The subject concerned had been excluded from the CORGI genome-wide association study because ancestry was uncertain in principal-component analysis. On further investigation, we found that the family of this individual was of Ashkenazi descent and had a phenotype of multiple polyps and CRC, entirely compatible with HMPS but not previously diagnosed as such (**Supplementary Fig. 2**). We found no non-Ashkenazi affected individuals with duplications in the region implicated in HMPS.

Although the *SCG5-GREM1* duplication was highly likely to be pathogenic and was an unusual type of mutation consistent with the seeming monophyletic origin of HMPS, we wished to exclude the presence of other potentially pathogenic variants on the HMPS haplotype. We therefore screened the entire ancestral haplotype region (chr. 15: 30,735,098–31,369,755) in 2 cases and 3 unaffected individuals

from 3 different families with HMPS through sequencing of 454 amplicons. In addition, we sequenced the region in two somatic cell hybrids that had been generated, one of which carried a copy of the mutant chromosome 15 and the other of which had a copy of the wild-type chromosome from an individual with HMPS. After filtering out variants that were present in dbSNP and a small number of artifacts, only three variants were specifically present in all cases and in no controls. All three variants were single-nucleotide changes deep within introns of *FMN1* (**Supplementary Table 2**). None of the changes was predicted to have profound functional effects, although one change at a conserved location at the beginning of *FMN1* (base 31,038,855) lay within a DNase I hypersensitive site, thus having the potential to affect gene transcription. Differences in *FMN1* transcription were subsequently excluded as a possible mechanism.

Given that the *SCG5-GREM1* duplication was by far the most likely pathogenic mutation in the HMPS region, we sought to determine which gene(s) was affected by this change. Because the duplication involved the latter half of the coding region of *SCG5* in addition to a large part of the region between *SCG5* and *GREM1*, we initially attempted to detect aberrant *SCG5* transcripts in the affected individuals on the basis of predictions from the duplicated exons. However, only the normal *SCG5* mRNA species was found (data not shown). We then wondered whether the duplication could act as an enhancer of gene transcription. mRNA levels of *SCG5*, *GREM1* and *FMN1* were assayed in 87 individual normal colonic crypts from 8 subjects with HMPS and in 55 individual crypts from 8 UK controls, as well as in 8 crypts from 1 family member with HMPS who did not have the duplication. The control subjects had clear colonoscopies and no history of colorectal tumors. We found markedly increased *GREM1* transcript levels in the

**Figure 2** Expression of *GREM1*, *SCG5* and *FMN1*. (**a**) Expression levels of each mRNA by quantitative RT-PCR (qRT-PCR) analysis of multiple individual crypts of normal morphology from subjects with HMPS relative to controls. Error bars, s.d. Note the greatly increased expression of *GREM1*. Immunohistochemistry provided no evidence of differences in *SCG5* expression in subjects with HMPS and controls (data not shown); reliable antibodies are not available for GREM1 and FMN1. (**b**) Representative light-(left) and dark-field (right) mRNA ISH of *GREM1* mRNA in colorectum of normal appearance from a control (unaffected) individual compared with an individual with HMPS. *GREM1* is restricted to ISEMFs and the subcryptal smooth muscle layer (indicated by arrows) in the control individual but is found throughout the crypt epithelium in the individual with HMPS.



normal epithelium of subjects with HMPS compared to the controls, but no significant differences in the expression of *SCG5* or *FMN1* were detected (**Fig. 2**). Allele-specific expression (ASE) analysis showed significantly increased expression of the duplicated allele in HMPS crypts (Mann-Whitney $P = 0.0009$). Attempts to confirm that the duplicated allele was expressed more strongly in lymphoblastoid cells from the subjects with HMPS were unsuccessful because we found *GREM1* levels to be very low in these cells. We did, however, examine expression in the somatic cell hybrids carrying the mutant or wild-type copy of chromosome 15, as these cells were derived from mouse fibroblast lines that expressed *GREM1*. We detected 4.6-fold higher *GREM1* mRNA expression in the cells with the mutant copy of human chromosome 15 compared to those carrying the wild-type copy.

We then examined the localization of *GREM1* mRNA in the normal colorectal crypts of subjects with HMPS and controls. In controls and family members with HMPS without the duplication, *GREM1* mRNA was restricted to intestinal subepithelial myofibroblasts (ISEMFs) at the crypt base, as has previously been reported[9,10]. In subjects with HMPS, by contrast, *GREM1* was expressed, not only in basal ISEMFs, but also at very high levels in epithelial cells (predominantly colonocytes), with expression extending most of the way up the sides of the crypt (**Fig. 2**). Increased *GREM1* expression was also seen in HMPS polyps, albeit to a lesser extent than in normal epithelium.

Given the increased and ectopic expression of *GREM1* in individuals with HMPS, we tested whether elements within the HMPS duplication could enhance gene transcription. Although data from the Encyclopedia of DNA Elements (ENCODE) suggest that at least two regions within the duplication might have enhancer activity (**Supplementary Fig. 3**), we focused on a 3-kb region (chr. 15: 30,779,000–30,782,000) because we had previously shown this to have histone H4 lysine 4 (H4K4) acetylation and histone H3 lysine 4 (H3K4) methylation marks indicative of active chromatin[11], and we
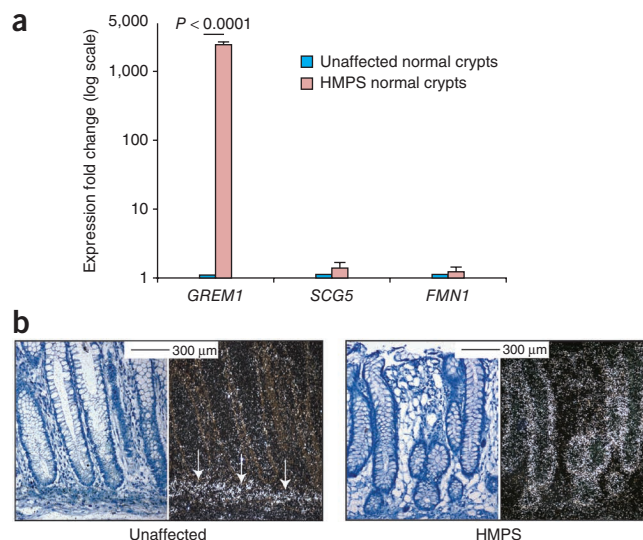
subsequently showed by chromatin immunoprecipitation that this region contained transcription factor–binding sites (M.B., A.L. and I.T., unpublished data). Using a luciferase reporter assay, we found that a fragment encompassing this 3-kb region was capable of enhancing gene expression by fourfold in the SW948 CRC cell line that expresses *GREM1* ($P = 0.00002$, *t* test; **Supplementary Table 3**). We then used chromatin conformation capture (3C) to show that the 3-kb region interacts directly with the *GREM1* promoter in *GREM1*-expressing CRC cell lines (data not shown). Using 3C analysis in the chromosome 15 somatic cell hybrids, we showed much stronger promoter interaction in the hybrid with the duplication than in the control (**Fig. 3**).

The BMP signaling pathway is not well characterized compared to other pathways. The canonical pathway signals through BMP receptors and drives expression of the ID transcriptional repressor genes through the activities of SMAD1, SMAD5 and SMAD8 (SMAD1/5/8). A non-canonical pathway acts through decreased PTEN expression and increased levels of active Akt[12]. BMP2 and BMP4, the primary ligands in the colorectum, have been postulated to act against the stem cell phenotype, and, hence, increased GREM1 levels would be expected to reduce BMP ligand levels and promote the stem cell phenotype. In keeping with this hypothesis, we found a small but significant increase in the expression of the stem and/or progenitor cell marker *LGR5* in whole HMPS colonic crypts compared to crypts from controls ($P = 0.03$, Mann-Whitney test). There was no detectable difference in staining for phosphorylated SMAD1/5/8 in lesions from
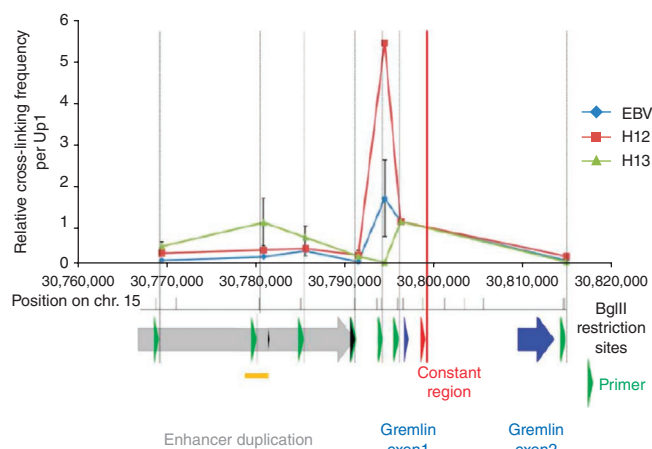


**Figure 3** 3C analysis of a region upstream of *GREM1* in somatic cell hybrids carrying a wild-type chromosome 15 or a copy with the HMPS duplication. The genomic region of *GREM1* is depicted below the *x* axis, with the telomeric limit of the duplication shown in gray and exons shown in blue. The 3-kb region within the duplication on which we focused is indicated by an orange bar. The 3C constant region and its primer and probe sites (red) are located on a BglII fragment containing the transcription start site and exon 1 of *GREM1*. Assayed BglII restriction fragments are shown as vertical gray lines. Variable human-specific primers (green) are located at the 3′ end of selected fragments. The relative cross-linking frequencies of each upstream fragment derived from qRT-PCR were normalized to a BAC control and calculated as the ratio of individual site cross-linking frequencies to the cross-linking frequency of the first upstream adjacent BglII fragment (Up1). The H13 hybrid cell line with the HMPS duplication showed a specific increase in the interaction of the 3-kb region with the *GREM1* promoter. The H12 cell line, containing the wild-type human allele, also showed some interaction between these sequences compared with the corresponding EBV-derived cell lines (EBV) that did not express *GREM1*.

subjects with HMPS, and we were not able to show an increase in ID1, ID2 and ID3 (ID1/2/3) expression in crypts from subjects with HMPS relative to controls. We did, however, detect decreased PTEN expression in some polyps from subjects with HMPS (but not in tissue of normal appearance), together with patches of complete PTEN loss (**Supplementary Fig. 4**), a finding of note, given the presence of hamartomatous polyps in the intestines of some individuals with Cowden syndrome who carry germline *PTEN* mutations[13]. There was a small increase in the levels of cytoplasmic phosphorylated Akt in polyps from subjects with HMPS (**Supplementary Fig. 4**), although this was not spatially identical to the changes in *PTEN* expression.

In summary, we have shown that the Mendelian colorectal polyposis syndrome HMPS results from a duplication of approximately 40 kb upstream of the gene that encodes the secreted BMP antagonist GREM1. Although the duplication includes some sequence within the neighboring *SCG5* gene, no difference in *SCG5* expression was found relative to controls. Whereas the duplication predisposes to benign lesions that are hypothesized to progress to CRCs in some cases, some sporadic adenomas and CRCs acquire low-level ectopic epithelial expression of *GREM1* (S.L., H.D., P.R.C., A.L. and I.T., unpublished data), suggesting that the increased levels of *GREM1* in HMPS may act to promote multiple stages of colorectal tumorigenesis. To date, we have only found the HMPS duplication in the Ashkenazi Jewish descendants of a single founder, but we do not exclude the possibility that rare affected individuals of other ancestry might carry pathogenic copy-number changes or other mutations in *GREM1* that modestly or considerably increase the risk of CRC. In this regard, it is noteworthy that a single CRC case with whole-gene duplication of *GREM1* has been reported[14], although, if pathogenic rather than incidental, this mutation produces features of hereditary non-syndromic CRC[15] rather than HMPS. It is also notable that our genome-wide association studies have shown that common SNPs (rs16969681 and rs11632715) within the HMPS duplication are associated with relatively modest differences in CRC risk in the general population[11]. We suggest that the HMPS duplication is causally associated with increased ectopic *GREM1* transcription in colorectal epithelium. Deletion of a limb bud–specific *Grem1* control region within *Fmn1* is known to lead to a limb deformity phenotype in mice[16]; we wonder whether the HMPS duplication involves an intestine-specific *GREM1* control region that lies partly within *SCG5*. Although no reliable antibody to GREM1 exists, it is expected that increased mRNA causes increased secretion of epithelial GREM1 protein and, hence, a reduction in BMP signaling to the crypt cells. A reduction in BMP signaling is also the likely cause of JPS, in which either the type 1A BMP receptor *BMPR1A* or the downstream BMP effector *SMAD4* is mutated and functionally defective[17,18]. The detailed functional consequences of the HMPS duplication, including full mapping of regulatory sites and the downstream effects of inactivating BMP signaling, await further investigation in the context of HMPS, but plausible modes of action include aberrant stem cell numbers or function and effects on the non-canonical BMP pathway via PTEN. It seems highly unlikely that all the cells overexpressing *GREM1* in the normal bowel of individuals with HMPS represent early tumor clones. One potential pathogenic mechanism underlying HMPS is an increase in the number of cells that are susceptible to tumor-causing mutations, probably through microenvironmental effects of increased *GREM1* secretion.

**URLs.** Integrated Regulation from ENCODE Tracks, http://genome. ucsc.edu/cgi-bin/hgTrackUi?hgsid=219213609&c=chr15&g= wgEncodeReg; Oxford Gene Technology aCGH, http://www.ogt.co.uk/ gs-arrayCGH.asp.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary information is available in the online version of the paper.*

### AUTHOR CONTRIBUTIONS
E.J., A.L., S.S., M.B., P.R.C., H.D., K.K., K. Heinimann, K. Howarth and S.L. performed laboratory experiments and analyzed data. J.T., S.L. and I.T. supervised laboratory experiments. H.T., J.E., S.L. and I.T. obtained samples. I.T. wrote the manuscript, with assistance from S.L. H.T. and I.T. oversaw the study.

### COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Whitelaw, S.C. *et al.* Clinical and molecular features of the hereditary mixed polyposis syndrome. *Gastroenterology* **112**, 327–334 (1997).
2. Jaeger, E.E. *et al.* An ancestral Ashkenazi haplotype at the *HMPS/CRAC1* locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am. J. Hum. Genet.* **72**, 1261–1267 (2003).
3. Tomlinson, I. *et al.* Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology* **116**, 789–795 (1999).
4. Jaeger, E. *et al.* Common genetic variants at the *CRAC1* (*HMPS*) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.* **40**, 26–28 (2008).
5. Cheah, P.Y. *et al.* Germline bone morphogenesis protein receptor 1A mutation causes colorectal tumorigenesis in hereditary mixed polyposis syndrome. *Am. J. Gastroenterol.* **104**, 3027–3033 (2009).
6. East, J.E., Saunders, B.P. & Jass, J.R. Sporadic and syndromic hyperplastic polyps and serrated adenomas of the colon: classification, molecular genetics, natural history, and clinical management. *Gastroenterol. Clin. North Am.* **37**, 25–46, v (2008).
7. Colella, S. *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **35**, 2013–2025 (2007).
8. Tomlinson, I.P. *et al.* COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br. J. Cancer* **102**, 447–454 (2010).
9. Kosinski, C. *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc. Natl. Acad. Sci. USA* **104**, 15418–15423 (2007).
10. Segditsas, S. *et al.* Putative direct and indirect Wnt targets identified through consistent gene expression changes in *APC*-mutant intestinal adenomas from humans and mice. *Hum. Mol. Genet.* **17**, 3864–3875 (2008).

11. Tomlinson, I.P. *et al.* Multiple common susceptibility variants near BMP pathway loci *GREM1*, *BMP4*, and *BMP2* explain part of the missing heritability of colorectal cancer. *PLoS Genet.* **7**, e1002105 (2011).

12. Tian, Q., He, X.C., Hood, L. & Li, L. Bridging the BMP and Wnt pathways by PI3 kinase/Akt and 14-3-3ζ. *Cell Cycle* **4**, 215–216 (2005).

13. Hobert, J.A. & Eng, C. *PTEN* hamartoma tumor syndrome: an overview. *Genet. Med.* **11**, 687–694 (2009).

14. Venkatachalam, R. *et al.* Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int. J. Cancer* **129**, 1635–1642 (2011).

15. Papaemmanuil, E. *et al.* Deciphering the genetics of hereditary non-syndromic colorectal cancer. *Eur. J. Hum. Genet.* **16**, 1477–1486 (2008).

16. Zuniga, A. *et al.* Mouse *limb deformity* mutations disrupt a global control region within the large regulatory landscape required for *Gremlin* expression. *Genes Dev.* **18**, 1553–1564 (2004).

17. Howe, J.R. *et al.* Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat. Genet.* **28**, 184–187 (2001).

18. Howe, J.R. *et al.* Mutations in the *SMAD4/DPC4* gene in juvenile polyposis. *Science* **280**, 1086–1088 (1998).

## ONLINE METHODS

**Subjects and samples.** Individuals and families were originally classified as possibly having a diagnosis of HMPS on the basis of the presence of multiple early-onset colorectal polyps (including some of serrated or mixed morphology) coupled with no obvious excess of the extra-colonic tumors typical of other Mendelian CRC syndromes. Samples having germline mutations in *APC*, *MUYTH*, *MSH2*, *MLH1*, *MSH6*, *PMS2*, *STK11* (also known as *LKB1*), *SMAD4* and *BMPR1A* were excluded. After identification of the Ashkenazi HMPS haplotype, the presence of this haplotype was used to confirm diagnosis, with this testing indicating that all Ashkenazi families with HMPS showed evidence of dominant inheritance of the disease. Any non-Jewish affected individuals (or Jewish affected individuals without the HMPS haplotype) with features suggestive of HMPS were classified as possibly having HMPS, although very few of these had evidence of dominantly inherited disease, and many were given a presumptive primary diagnosis of hyperplastic polyposis syndrome[6].

Ethical approval for the study was provided by the Southampton and South-West Hampshire Research Ethics Committee A. After obtaining informed consent, peripheral blood samples were taken, and DNA was extracted from these subjects. Tumor samples, including CRCs, were obtained from histo-pathology archives. Fresh biopsies of normal bowel were additionally taken from a number of individuals with HMPS at colonoscopy and from tumor-free individuals who were having investigative colonoscopy for symptoms subsequently shown to have innocent origins.

Lymphoblastoid cell lines were derived from several subjects with HMPS and unaffected family members and from one affected individual; two somatic cell hybrids were derived in a mouse NIH-3T3 background, with each hybrid containing a single copy of chromosome 15 carrying the mutant or wild-type haplotype at the HMPS locus.

**Genotyping and variant screening.** Microsatellite genotyping employed fluorescence-based fragment analysis on ABI3730 sequencers.

Copy-number analysis was undertaken using an Oxford Gene Technology (OGT) custom-designed oligonucleotide microarray comprising 6,761 probes spanning a 2.2-Mb region that was delimited on the basis of critical flanking recombinants from linkage analysis data available at the time. In brief, test DNA (from subjects with HMPS or unaffected family member) was labeled with Cy5, and control DNA (a standard reference) was labeled with Cy3. DNA was simultaneously hybridized to the arrays, and the ratios of Cy3/Cy5 intensities were measured. Copy-number gain was assessed as a consistent relative increase in Cy5, and copy-number loss was assessed as an increase in Cy3, using proprietary OGT software.

Illumina Infinium Hap550 arrays were used for SNP genotyping and subsequently for copy-number analysis. HMPS samples were analyzed alongside a larger sample set from the CORGI genome-wide association study of CRC susceptibility in the general UK population. Quality control analysis was performed as described for the CORGI study[11]. Copy-number analysis on data from the Hap550 arrays was performed using the QuantiSNP program[7].

DNA sequencing was originally performed using conventional Sanger sequencing on the ABI3730 sequencer, and analysis was performed by visual inspection of sequence electropherograms from the mutant and wild-type somatic cell hybrids, on the basis of human-specific amplicons. More recently, sequencing was performed using the Roche454 GSFlex system and software. The region of interest was tiled with long-range PCR amplimers, and amplification products were pooled at equal stoichiometric concentrations. Sequencing libraries were prepared using the Roche 454 Rapid Library preparation method, and each sample was tagged with a unique molecular identifier before pooling and emulsion PCR amplification. In total, 200 Mb of sequencing data were generated, which was demultiplexed and analyzed using Roche gsMapper software.

Information on all PCR primer sequences and reaction conditions are available on request. Human Genome Build36 was used as a reference throughout this manuscript.

**Quantitative RT-PCR.** The TaqMan system was used to analyze the expression of *GREM1*, *SCG5*, *FMN1* and *LGR5*. Small biopsies of normal bowel were taken from the ileum and throughout the colorectum of HMPS cases and controls. Tissue was incubated in 5 ml of DMEM (Invitrogen) supplemented with 30 mM EDTA (Sigma), 0.1 M dithiothreitol and 100 µl of RNAlater (Ambion) for 15 min. Tissue was transferred to PBS and then vigorously shaken for 20 s. Individual intestinal crypts were drawn up using glass pipettes and transferred to RNA lysis buffer (Qiagen). After several cycles of shaking with fresh PBS washes, complete epithelial denudement occurred, and the residual mesenchymal tissue was collected and processed. Extracted mRNA was analyzed by qRT-PCR on the ABI7700 system. Primers and probes for test mRNA (details on request) were obtained from the ABI catalog. A *GAPDH* control probe was used. Expression levels were determined using the standard $\Delta\Delta C_T$ method.

**Allele-specific expression (ASE).** ASE analysis was performed following a previously described protocol[19]. A TaqMan SNP genotyping assay for rs12915554 within the 3′ UTR of the *GREM1* gene was used to identify two subjects with HMPS and three control subjects that were heterozygotes. Phase was assigned using existing genotyping data, including those from the somatic cell hybrids. DNA from control homozygote was mixed in different ratios, and fluorescent intensity data were used to generate a linear regression standard curve. ASE for the heterozygotes with HMPS and control cDNA from individual crypts were then measured using qRT-PCR, and the gene expression allelic ratio was measured by intercepting the log of the fluorescent intensity ratio on the standard curve.

**In situ tissue analyses.** Sections (4 µm) were dewaxed and rehydrated by standard methods. Endogenous peroxidase activity was blocked with 3% $H_2O_2$ in methanol for 10 min. Antigen retrieval was achieved by 10 min of pressure cooking in sodium citrate buffer at pH 6. Slides were blocked in goat serum for 30 min. The following primary antibodies were used for immunohistochemistry: rabbit antibody to phosphorylated SMAD1/5/8, which recognizes the phosphorylated forms of SMAD1 (Ser463/465), SMAD 5 (Ser463/465) and SMAD8 (Ser426/428) (1:50 dilution; Cell Signaling Technology, 9511), rabbit antibody to PTEN (1:100 dilution; Cell Signaling Technology, 9188) and rabbit antibody to Akt phosphorylated at Ser473 (1:50 dilution; Cell Signaling Technology, 4060). Then, samples were incubated with a 1:200 dilution of biotinylated polyclonal goat antibody recognizing rabbit (DAKO, E0432) before incubation in ABC (containing avidin and biotinylated horseradish peroxidase) (Vector Labs). Each antibody was applied for 30 min, and three 5-min PBS washes were performed. Sections were then developed with 3,3-diaminobenzidine-tetrahydrochloride solution (DAB; Sigma) for 2 min, rinsed in tap water and stained with light hematoxylin counterstain. Negative controls underwent the same steps but were incubated with PBS instead of the primary antibody solution.

Because suitable antibodies continue to be unavailable, expression levels of *GREM1*, *FMN1* and *LGR5* were analyzed by mRNA ISH using a previously described method[20] with [35S]-labeled riboprobes. *SCG5* was primarily analyzed in this way, although immunohistochemistry was also used to provide supporting evidence. Details of the riboprobes used are available on request. Expression was scored in a qualitative fashion as present or absent or as stronger or weaker than in control samples analyzed alongside the samples from individuals with HMPS.

**Assessment of the control of GREM1 expression.** For luciferase reporter assays, the region chr. 15: 30,778,456–30,782,505 was cloned into the pGL3-Promoter vector (Promega) in forward and reverse orientations to test enhancer function. SW948 cells were transiently co-transfected (in triplicate) with the appropriate pGL3-Promoter constructs and the *Renilla* luciferase pGL4.75 vector (Promega), as a control for transfection efficiency. After 48 h, luciferase activity was measured (Dual-Glo Luciferase Assay System, Promega), and firefly luciferase activity from enhancer regions was normalized to *Renilla* luciferase values for each sample. Background levels of firefly luciferase activity were assayed by transfecting cells with a control plasmid consisting of a noncoding 2.2-kb stretch of plasmid sequence (taken from the pENTR1A plasmid, Invitrogen) cloned into the pGL3-Promoter vector. These values were further used to normalize test luciferase measurements.

3C was performed following a previously described protocol[21]. Approximately $1 \times 10^7$ cells (at 90–100% confluence) were cross-linked with 1% formaldehyde for 20 min and quenched with 125 mM glycine. After shaking the cells for more than 30 min in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100 and protease inhibitors),

nuclei were pelleted (300*g* for 5 min) and resuspended in 360 μl of H$_2$O and 60 μl of the 10× restriction buffer for BglII. After 1 h of incubation in 3.75% SDS and subsequent quenching with 37.5% Triton X-100, 400–800 U of BglII was added, and samples were incubated overnight at 37 °C. The ligation reaction used a 10× higher volume to favor ligation events between cross-linked DNA strands with 50 U of T4 Ligase (Roche, 5 U/μl), and reactions were incubated at 16 °C over 1–2 nights. To reverse cross-linking, 30 μl of Proteinase K (10 mg/ml) was added, and reactions were incubated overnight at 65 °C. After phenol extraction and ethanol precipitation, the samples were recovered in 500 μl of TrisEDTA buffer (Sigma). Elute (4 μl) was used for qRT-PCR in the TaqMan Fast Universal PCR System (Applied Biosystems).

Data were normalized to cross-linking frequencies for the RP11-420B6 BAC. The ratio of the test data points to the adjacent upstream primer data was used to compare results between cell lines.

19. Lo, H.S. *et al.* Allelic variation in gene expression is common in the human genome. *Genome Res.* **13**, 1855–1862 (2003).
20. Poulsom, R., Longcroft, J.M., Jeffery, R.E., Rogers, L.A. & Steel, J.H. A robust method for isotopic riboprobe *in situ* hybridisation to localise mRNAs in routine pathology specimens. *Eur. J. Histochem.* **42**, 121–132 (1998).
21. Miele, A., Gheldof, N., Tabuchi, T.M., Dostie, J. & Dekker, J. Mapping chromatin interactions by chromosome conformation capture. *Curr. Protoc. Mol. Biol.* **Chapter 21**, Unit 21.11 (2006).

npg