Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X

Eduard Schulz, Petra Klampfl, Stefanie Holzapfel, Andreas R. Janecke, Peter Ulz, Wilfried Renner, Karl Kashofer, Satoshi Nojima, Anita Leitner, Armin Zebisch, Albert Wölfler, Sybille Hofer, Armin Gerger, Sigurd Lax, Christine Beham-Schmid, Verena Steinke, Ellen Heitzer, Jochen B. Geigl, Christian Windpassinger, Gerald Hoefler, Michael R. Speicher, C. Richard Boland, Atsushi Kumanogoh, Satoshi Nojima, Karl Kashofer, Jochen B. Geigl, Christian Windpassinger, Gerald Hoefler, Michael R. Speicher, C. Richard Boland, Atsushi Kumanogoh, Gerald Hoefler, Michael R. Speicher and Heinz Sill

Familial colorectal cancer type X (FCCTX) is characterized by clinical features of hereditary non-polyposis colorectal cancer with a yet undefined genetic background. Here we identify the SEMA4A p.Val78Met germline mutation in an Austrian kindred with FCCTX, using an integrative genomics strategy. Compared with wild-type protein, SEMA4AV78M demonstrates significantly increased MAPK/Erk and PI3K/Akt signalling as well as cell cycle progression of SEMA4A-deficient HCT-116 colorectal cancer cells. In a cohort of 53 patients with FCCTX, we depict two further SEMA4A mutations, p.Gly484Ala and p.Ser326Phe and the single-nucleotide polymorphism (SNP) p.Pro682Ser. This SNP is highly associated with the FCCTX phenotype exhibiting increased risk for colorectal cancer (OR 6.79, 95% CI 2.63 to 17.52). Our study shows previously unidentified germline variants in SEMA4A predisposing to FCCTX, which has implications for surveillance strategies of patients and their families.
Colorectal cancer (CRC) is the third most common cancer worldwide\(^1\). Approximately 5% of cases are inherited in an autosomal dominant manner with familial adenomatous polyposis and hereditary non-polyposis colorectal cancer (HNPPC) being the two major hereditary forms\(^2,3\). HNPPC is clinically diagnosed when Amsterdam-I or -II criteria (AC-I/II) are met: three or more relatives affected through at least two generations by CRC (AC-I) or an HNPCC-associated cancer (AC-II), respectively, with one patient being a first-degree relative of the other two and one diagnosed before the age of 50 years\(^4\). However, 40 to 50% of patients with HNPPC fulfilling AC-I lack detectable germline mutations in cancer predisposition genes and are classified as familial colorectal cancer type X (FCCTX)\(^5,6\). In contrast to Lynch syndrome (LS)—the HNPPC entity characterized by germline DNA mismatch repair (MMR) gene mutations and somatically acquired microsatellite instability—individuals with FCCTX exhibit decreased risk for extracolonic neoplasms, that is, endometrial, stomach, small bowel and urinary tract carcinomas and tumour formation including CRC. Development tends to occur at a later age\(^5,8,9\). It is expected that single uncommon susceptibility genes transmitted in an autosomal dominant manner are responsible for a subset of FCCTX cases, which in turn implies that this syndrome is likely to be heterogeneous\(^2,5,8\). Here we show that germline variants in the semaphorine 4A (SEMA4A) gene confer susceptibility to FCCTX. This finding broadens our understanding of the biology of those malignancies and forms the basis for effective cancer detection and prevention strategies.

### Results

**Pedigree analysis and variant identification.** In the course of a previous study focusing on pedigree analysis of patients with therapy-related myeloid neoplasms\(^10,11\), we have identified a large Austrian kindred with FCCTX (Family K, Fig. 1a; Supplementary Fig. 1). CRCs in this family were inherited in an autosomal dominant pattern with incomplete penetrance meeting AC-I. In each affected individual, one to six colorectal adenomas and one to two CRCs were diagnosed at a median age of 62.5 years (range, 44–72). The majority of colorectal neoplasms was located in the distal colon and rectum and showed tubular histological features without evidence for an increase of infiltrating lymphocytes (Table 1).

We conducted genetic linkage analysis (LA) of five family members with colorectal neoplasms and one unaffected, putative mutation carrier (Fig. 1a), which revealed four shared regions on chromosomes 1, 3, 10 and 20 (Supplementary Fig. 2), none of them harbouring known cancer-associated genes. We next performed whole-exome sequencing (WES) on four of these individuals (Fig. 1a). A heterozygous germline variant was identified in the MUTYH gene (NM_001128425.1:c.650G>A: p.Arg217His, rs147754007) in the first-degree relatives K13 and K18 but not in individuals K3 and K14 (Supplementary Fig. 3). We, therefore, excluded MUTYH R217H as a culprit germline mutation responsible for the majority of neoplasms in this family, which is in line with the fact that MUTYH-associated polyposis is an autosomal recessive CRC predisposition syndrome\(^12\). To identify novel candidate causative mutations, we combined LA and WES and filtered heterozygous, non-synonymous protein-coding or splice-site variants with a minor allele frequency of \(<0.01\) (Supplementary Table 1). All variants were confirmed by Sanger sequencing and analysed in two further family members with CRC (K16 and K26). Only variant p.Val78Met (NM_001193300:c.232G>A) in the SEMA4A gene located on chromosome 1q22 was shared by all tested individuals. However, in this approach, we included two individuals with colorectal adenomas constituting a frequent but not obligate part of HNPPC syndromes\(^13,14\). As this might constitute a potential bias, we focused in an independent analysis on variants from WES shared by individuals with CRC (K13, K18) or with an offspring with CRC (K3). Of 24 variants identified (Supplementary Table 2), two were also present in individuals K16 and K26. We excluded the p.Val212Phe variant in ZNF763 (rs7249379) due to non-conservation because Phe212 represents the common chimpanzee allele. Only SEMA4A V78M segregated with all CRC cases and was also detected in individuals K9 with testicular and K14 with breast cancer, respectively (Fig. 1a). Given a mean age of 61 years of individuals with FCCTX at disease onset\(^1\), we estimated a phenocopy rate of 0.00 and a penetrance rate of 0.56 of the SEMA4A V78M variant in Family K. cDNA from peripheral blood (PB) leukocytes demonstrated expression of the mutant allele (Supplementary Fig. 4).

SEMA4A is a membrane-bound class 4 semaphorin receptor with organ-specific and immunomodulatory effects as well as growth regulatory functions\(^14–16\). V78M lies within the SEMA domain responsible for receptor binding and Val78 is well conserved (Fig. 2a; Supplementary Fig. 5). This variant is absent from dbSNP137, the 1000 Genomes Project database and the National Heart, Lung and Blood Institute Exome Variant Server (ESP6500). Prediction tools favour consequences for its protein function (SIFT score = 0, PolyPhen-2 score = 0.987, vertebrate PhyloP100 score = 7.434, vertebrate PhastCons100 score = 1, phastConsElements100 score = 407 [LOD = 65] and MutationTaster 2 = disease causing with 0.95 probability value).

### Recurrent somatic mutations in CRCs of SEMA4A V78M carriers.

We then analysed CRC specimens of mutation carriers for copy-number alterations by array-based comparative genomic hybridization and loss of heterozygosity (LOH) by Sanger sequencing, respectively. Gains on the long arm of chromosome 1 involving the SEMA4A locus were observed in two of three CRCs together with a homozygous SEMA4A V78M status (Fig. 3). We did not detect copy-number alterations in the MUTYH gene in any of the three analysed CRCs including the heterozygous R217H carrier K13. We also analysed four available CRCs for recurrent, somatically acquired mutations in known CRC genes by targeted deep sequencing and identified mutations in TP53 in 3/4, APC in 2/4, KRAS in 2/4 and PIK3CA in 1/4 CRC cases, respectively, as possible cooperating events (Table 2). Notably, there was no predominance of C:G to A:T transversion mutations in the CRC of patient K13 characteristic for complete loss of MUTYH activity\(^12\).

**SEMA4A V78M affects proliferative pathways.** Compound heterozygous germline mutations in SEMA4A have been reported in patients with retinal degenerative diseases and studies in knock-in mice showed that one of these mutations (F350C) leads to an abnormal Sem4A4A localization in retinal pigment epithelial cells\(^17,18\). A three-dimensional protein model of human SEMA4A predicts that Val78 has no spatial relationship to residues associated with retinal disorders (Fig. 2b). In agreement with this prediction and the family’s history lacking apparent ocular manifestations, the expression of a fusion gene composed of Sem4A4A V78M and carboxyl-terminal green fluorescent protein (GFP) in human retinal ARPE-19 cells showed normal GFP signal distribution (Fig. 4a).

SEMA4A is widely expressed including normal colonic tissue (Supplementary Fig. 6) but is undetectable in 2/4 CRC cell lines analysed (Supplementary Fig. 7). It has been shown to have inhibitory effects on proliferation and migration of endothelial cells by antagonizing vascular endothelial growth factor\(^16\).
We therefore analysed transiently transfected SEMA4A-deficient HCT-116 cells characterized by KRAS and PIK3CA mutations. We were unable to demonstrate significant differences between wild-type and mutant SEMA4A on migration (Supplementary Fig. 8). However, as compared with SEMA4A wt, significantly more SEMA4AV78M-transfected cells were in S phase under normal growth conditions (Fig. 4b,c). We then assessed activation of the phosphoinositide 3-kinase/Akt (PI3K/Akt), mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/Erk) and Wnt/b-catenin pathways that have been shown to be important in colorectal carcinogenesis 19. As compared with SEMA4A wt, SEMA4A V78M-transfected HCT-116 cells revealed significantly enhanced activation of the PI3K/Akt and MAPK/Erk pathways both mediating proliferation by increasing cells in S phase and accelerating G2/M transition (Fig. 4d,e; Supplementary Fig. 9) 20–22. Transient transfection of 293T cells, however, showed no effect of SEMA4A on the PI3K/Akt pathway (Supplementary Fig. 10).

SEMA4A variants are associated with FCCTX. To study the prevalence of SEMA4A germline mutations in FCCTX, we screened 53 unrelated FCCTX cases from Austria, Germany and the United States (Supplementary Table 3) and identified two further mutations located in the SEMA domain (heterozygous c.1451G>T, p.Gly484Ala, rs148744804; homozygous c.977C>T, p.Ser326Phe; Supplementary Fig. 11). These mutations affect highly conserved residues (Fig. 2a and Supplementary Figs 12 and 13) and prediction tools indicate an effect on protein function for both of them (Supplementary Table 4). The G484A variant has a global minor allele frequency of 0.001 in the 1000 Genomes Project and ESP6500 databases. It was also found in the index patient’s brother affected with CRC (Fig. 1b; Supplementary Fig. 11). The novel S326F variant affects a residue predicted to be involved in homodimer formation (Fig. 2b; Supplementary Fig. 12). Furthermore, we detected the heterozygous single-nucleotide polymorphism (SNP) p.Pro682Ser (c.2044C>T, rs76381440) in six of 47 (13%) German and Austrian FCCTX patients, respectively (Supplementary Table 3; Supplementary Fig. 11). We, therefore, initiated a genetic association study using DNA from 1,138 Caucasian control subjects from Austria without a personal or family history of cancer. These specimens were collected previously during the course of a local health screening study23. The P682S SNP demonstrated a highly significant association with the FCCTX phenotype resulting in an increased risk for CRC (Table 3). Screening the 1000 Genomes Project data base revealed a comparable prevalence of heterozygotes among European individuals of 2.0%.
SEMA4A is somatically mutated in sporadic cancers. Finally, we were interested whether somatically acquired SEMA4A mutations are prevalent in sporadic CRCs as well as other neoplasms. Analysis of confirmed mutations across different cancer types revealed that SEMA4A mutations occur in 2.7% (15/559) of colorectal, 2.8% (6/212) of stomach and 3.3% (8/241) of uterine cancers24,25. In 92% of them, they constitute missense mutations (Supplementary Table 5) scattered throughout the gene (Fig. 2a).

Data from the cBioPortal for Cancer Genomics indicate that the SEMA4A gene is amplified in a wide range of different tumours and that deletions are only rarely seen (Supplementary Fig. 14).

**Table 1 | Clinical characteristics of colorectal neoplasms of Family K exhibiting the germline V78M SEMA4A mutation.**

| Patient | Neoplasm | Age (years) | Histology | Grading/staging | Localization | SEMA4A V78M |
|---------|----------|-------------|-----------|-----------------|--------------|-------------|
| K6      | CRA      | 44          | Tubular adenoma | Well to moderately differentiated | NA | + |
| K6      | CRA      | 61          | Tubular adenoma | Well differentiated | Sigmoid colon | + |
| K10     | CRA      | 66          | Tubular adenoma | Well to moderately differentiated | NA | + |
| K13     | CRC      | 48          | Adenocarcinoma | pG-3, pT-4, pN-1 | Coecum | + |
| K14     | CRA      | 63          | Tubular adenoma | Well differentiated | Rectum | + |
| K16     | CRC      | 71          | Tubulopapillary and mucinous adenocarcinoma | pG-2, pT-2, N-0 | Coecum | + |
| K16     | CRA      | 71          | Tubulovillous adenoma | Well to moderately differentiated | Coecum | + |
| K16     | CRC      | 72          | Tubular adenocarcinoma | pG-2, pT-X | Descending/sigmoid colon | + |
| K17     | CRA      | 46          | Tubular adenoma | Well differentiated | Rectum | – |
| K18     | CRC      | 62          | Tubular adenocarcinoma | pG-2, pT-1, N-0 | Sigmoid colon | + |
| K18     | CRA      | 62          | Tubular adenoma | Well to moderately differentiated | Sigmoid colon | + |
| K18     | CRC      | 62          | Tubular adenoma | Well to moderately differentiated | Rectum | – |
| K18     | CRA      | 64          | Tubular adenoma | Well to moderately differentiated | Ascending colon | + |
| K18     | CRA      | 65          | Tubular adenoma | Well to moderately differentiated | Descending colon | + |
| K18     | CRM      | 66          | Tubular adenoma | Well to moderately differentiated | NA | + |
| K18     | CRA      | 67          | Tubular adenoma | Well to moderately differentiated | Descending colon | + |
| K20     | CRM      | 55          | Tubulopapillary adenoma | Well differentiated | Sigmoid colon | + |
| K26     | CRC      | 55          | Adenocarcinoma | pG-2, pT-3, pN-2 | Rectum | + |

CRA, colorectal adenoma; CRC, colorectal cancer; NA, not available.

**Figure 2 | Localization of germline and somatic CRC SEMA4A mutations at the protein level.** (a) Germline mutations found in this study are illustrated in red, the SNP in orange, germline mutations associated with eye diseases in blue and somatic CRC mutations in black, respectively. Multiple sequence alignments of SEMA4As of selected species are shown below. Note that class 4 semaphorins can only be found in vertebrates. (b) SEMA and PSI domains (55–527, yellow) of human SEMA4A were modelled primarily to SEMA4D (1OLZ). Eye disease-associated residues D345 and F350 are located in the back of the protein below the plexin binding sites (magenta). V78 and G484 have no contact to the surface; are spatially distinct from D345 and F350 but are located in juxtaposition in β-propellers 1 and 7, respectively. S326 is part of the homodimer interface (cyan) having surface contact.

SEMA4A is somatically mutated in sporadic cancers. Finally, we were interested whether somatically acquired SEMA4A mutations are prevalent in sporadic CRCs as well as other neoplasms. Analysis of confirmed mutations across different cancer types revealed that SEMA4A mutations occur in 2.7% (15/559) of colorectal, 2.8% (6/212) of stomach and 3.3% (8/241) of uterine cancers24,25. In 92% of them, they constitute missense mutations (Supplementary Table 5) scattered throughout the gene (Fig. 2a).

Data from the cBioPortal for Cancer Genomics indicate that the SEMA4A gene is amplified in a wide range of different tumours and that deletions are only rarely seen (Supplementary Fig. 14).

**Discussion**
Semaphorins constitute a family of secretory or membrane-bound receptors, which were first described as regulators of neuronal axon growth26. They are characterized by an extracytoplasmic amino-terminal β-propeller—the SEMA
Figure 3 | LOH in CRCs of patients K16 and K26. (a) Array-based comparative genomic hybridization of three CRCs from Family K with germline SEMA4A V78M mutation. A gain in the SEMA4A locus is marked with an arrow. (b) Sanger sequencing. (c) Quantitative dPCR using fluorophore-coupled (VIC, FAM) TaqMan probes specific for wild-type (V78) or mutant (M78) SEMA4A nucleotide variants. Each dot represents a single well on a 20K chip. The performance of this assay was tested with specific oligonucleotide templates. The confidence level was set to 95% and the desired precision value was 10%. NTC, no template control.
The compound heterozygous germline SEMA4A variants D345H and F350C have been described in patients with retinitis pigmentosa and cone rod dystrophy but until now this finding has not been replicated. Semaphorin 4A-deficient mice exhibit photoreceptor degeneration and disturbed T-helper cell function but lack apparently increased tumour development. Given the wide expression of SEMA4A in different tissues, it is plausible that mutations can have different effects depending on the respective tissue. In fact, only the F350C but not the D345H variant was able to recapitulate the retinal disease phenotype of Semaphorin 4A-deficient mice in a homozygous knock-in mouse model, a genotype not described in humans yet. This observation stresses the special role of the F350 residue for photoreceptor function. The fact that these mice do not develop overt tumours does not necessarily argue against a potential tumour predisposing role. First, these animals have not been thoroughly investigated for tumour formation, and second, mutations in human cancer susceptibility gene homologues do not consistently result in increased carcinogenesis in mice. With respect to colorectal carcinogenesis, this has been clearly shown for the MMR gene Pms2 as well as for Smad4 predisposing to LS and juvenile polyposis syndrome, respectively. For both conditions, additional germline truncating mutations in the gatekeeper gene Apc are needed for intestinal tumour development in mice.

SEMA4A variants found in this study were not restricted to a certain hot spot region indicating a loss-of-function mechanism. This assumption is further supported by functional in vitro assays performed in the SEMA4A-deficient CRC cell line HCT-116. Whereas activation of mitogenic pathways like MAPK/Erk and PI3K/Akt within these cells could be diminished by transfection of a SEMA4A WT construct, expression of the SEMA4A V78M mutant failed to do so. Accordingly, re-expression of SEMA4A WT but not SEMA4A V78M inhibited G2/M-phase transition in HCT-116, again suggesting a loss-of-function of the V78M substitution. It has to be mentioned that the results of our copy-number analysis demonstrated a gain of chromosome 1q22; however, homozygosity of the V78M variant observed in two of the CRCs could nevertheless indicate that SEMA4A acts as a tumour suppressor rather than a proto-oncogene in the context of familial colorectal tumorigenesis. Middeldorp et al. found that tumour specimens from patients with FCCTX frequently exhibit gains of different chromosomal regions including chromosome 1, which is accompanied by copy-neutral LOH. Loss of the SEMA4A wild-type allele accompanied by amplification of the mutant one might be one mechanism of tumour suppressor inactivation in this particular entity. Unfortunately, due to low-quality DNA obtained from formalin-fixed, paraffin-embedded (FFPE) tumour specimens as well as lack of appropriate heterozygous microsatellite loci within or adjacent to the SEMA4A gene, we were unable to prove the type of LOH in tumours of Family K. Whether public data indicating that the SEMA4A gene is predominantly amplified in diverse cancers can also be additional germline mutation unrelated to other familial cancer cases or may indicate an autosomal recessive mode of inheritance operational in this family. However, due to lack of DNAs from other family members, we were unable to resolve this issue.

The SEMA4A P682S SNP is associated with an increased risk of CRC in our association study including Austrian and German individuals. Although this finding has to be replicated in an independent cohort and might reveal ethnic differences, the data, nevertheless, suggest that P682S constitutes a risk allele for a small proportion of CRC cases probably missed by genome-wide association studies that detect mostly frequent, low penetrant susceptibility loci.

The germ line variants found in this study were not restricted to a certain hot spot region indicating a loss-of-function mechanism. This assumption is further supported by functional in vitro assays performed in the SEMA4A-deficient CRC cell line HCT-116. Whereas activation of mitogenic pathways like MAPK/Erk and PI3K/Akt within these cells could be diminished by transfection of a SEMA4A WT construct, expression of the SEMA4A V78M mutant failed to do so. Accordingly, re-expression of SEMA4A WT but not SEMA4A V78M inhibited G2/M-phase transition in HCT-116, again suggesting a loss-of-function of the V78M substitution. It has to be mentioned that the results of our copy-number analysis demonstrated a gain of chromosome 1q22; however, homozygosity of the V78M variant observed in two of the CRCs could nevertheless indicate that SEMA4A acts as a tumour suppressor rather than a proto-oncogene in the context of familial colorectal tumorigenesis. Middeldorp et al. found that tumour specimens from patients with FCCTX frequently exhibit gains of different chromosomal regions including chromosome 1, which is accompanied by copy-neutral LOH. Loss of the SEMA4A wild-type allele accompanied by amplification of the mutant one might be one mechanism of tumour suppressor inactivation in this particular entity. Unfortunately, due to low-quality DNA obtained from formalin-fixed, paraffin-embedded (FFPE) tumour specimens as well as lack of appropriate heterozygous microsatellite loci within or adjacent to the SEMA4A gene, we were unable to prove the type of LOH in tumours of Family K. Whether public data indicating that the SEMA4A gene is predominantly amplified in diverse cancers can also be additional germline mutation unrelated to other familial cancer cases or may indicate an autosomal recessive mode of inheritance operational in this family. However, due to lack of DNAs from other family members, we were unable to resolve this issue.

The SEMA4A P682S SNP is associated with an increased risk of CRC in our association study including Austrian and German individuals. Although this finding has to be replicated in an independent cohort and might reveal ethnic differences, the data, nevertheless, suggest that P682S constitutes a risk allele for a small proportion of CRC cases probably missed by genome-wide association studies that detect mostly frequent, low penetrant susceptibility loci.

| Table 2 | Results of targeted deep sequencing of cancer hot spot regions in CRCs from Family K with the germline V78M SEMA4A mutation. |
|----------|---------------------------------------------------------------|-------------------|
| Patient  | Somatic mutation                                         | Protein alteration | dbSNP141        |
| K13      | APC NM_000385.5:c.2626C>T                                 | p.R876X           | rs12191333      |
| K13      | APC NM_000385.5:c.4348C>T                                 | p.R1450X          | rs12191332      |
| K13      | KRAS NM_004985.4:c.34_35delinsAT                           | p.G12t            | NA               |
| K13      | KRAS NM_004985.4:c.380C>T                                 | p.S127T           | NA               |
| K16      | None found                                              | None found        | —                |
| K18      | TP53 NM_000546.5:c.844C>T                                 | p.R282W           | NA               |
| K26      | APC NM_000385.5:c.4135G>T                                 | p.E1379X          | rs12191332      |
| K26      | KRAS NM_004985.4:c.34G                                     | p.G12S            | NA               |
| K26      | TP53                                                     | p.R248Q           | rs11540682      |
| NM_000546.5:c.743G>A |                                  |                  |                  |

CRC, colorectal cancer; NA, not available.
Figure 4 | SEMA4AV78M shows normal surface expression and leads to cell cycle changes in HCT-116 cells. (a) ARPE-19 cells were transfected with the plasmid constructs expressing Sema4A\textsubscript{WT}-EGFP or Sema4A\textsubscript{V78M}-EGFP proteins, incubated for 48 h and stained with phalloidin. Green, Sema4A-EGFP; red, phalloidin (actin). Representative images obtained by confocal microscopy are shown. The size of the scale bar is 20 μm. (b,c) Representative density plots and statistical analysis of GFP-positive SEMA4A-transfected HCT-116 cells stained by 7-AAD and APC anti-BrdU antibodies for cell cycle analysis. Cells were analysed 48 h after transfection. Significantly, more SEMA4AV78M than SEMA4AWT-transfected cells are in S phase and significantly less in G2/M phase, respectively (mean ± s.e.m.; n = 3 per group; two-tailed paired Student’s t-test; *P < 0.05 compared with WT). Cell cycle phase: Sub-G1 (R1), G1/G0 (R2), S (R3), G2/M (R4). (d,e) Representative immunoblots and statistical analysis of SEMA4A-transfected HCT-116 cells (whole-cell lysates) lysed 48 h after transfection. SEMA4AV78M-transfected cells show increased phosphorylation of Akt and Erk (mean ± s.e.m.; n = 6 per group; two-tailed paired Student’s t-test; *P < 0.05 compared with WT). (p-)GSK3β and (active) β-catenin proteins were blotted on a separate membrane in this experiment. No effects on GSK3β and β-catenin phosphorylation were seen in repeated experiments.
interpreted this way, should be handled with extreme caution as context specific functions have to be taken into account. Indeed, it has recently been shown that solubilized Sema4A at high levels is able to suppress cell death induced by plexin D1 in the mouse mammary tumour cell line 4T1, whereas the identical constellation inhibited proliferation in human endothelial cells.16,39

Our in vitro results have shown that Sema4A differentially modulates the PI3K/Akt and MAPK/Erk pathways in HCT-116 cells and that additional molecular hits are likely needed to establish the Sema4A phenotype, which is in accordance with well-established concepts of predisposing germline mutations.37 For instance, mutations in the Pik3ca and Kras genes found in the HCT-116 cell line could represent additional oncogenic hits. Recently, two different molecular entities have been postulated among FCCTX families with respect to somatically acquired aberrations found in their CRCs. One entity exhibiting loss of tumour suppressor loci involving the Tp53, ApC, Smad4 and Dcc genes as well as mutations in ApC and kras and another one with stable genotypes at these loci.40-42 Although our data demonstrating somatic mutations in the Tp53, ApC and Kras genes in CRCs from Family K are in line with these results, the numbers of tumours studied are too small to draw a final conclusion especially with respect to cooperation with Sema4A.17,38

In summary, the data presented here broaden our understanding of the pathophysiological role of semaphorins in human carcinogenesis and will have important consequences for screening and early tumour detection strategies of patients with FCCTX and their family members.

Methods

Subjects and primary samples. The study was approved by the institutional review board of the Medical University of Graz, Graz, Austria (MUG) and conducted according to the declaration of Helsinki. Written informed consent was obtained from each study participant or, in the case of deceased patients, close relatives for providing personal and family history data as well as biological specimens. Some of them were processed and stored by the Biobank of MUG.

Family K (germline Sema4A p.Val78Met) was from southern Austria and consisted of 88 members spread into two branches. Clinical data revealed that AC-I and/or AC-II criteria.13,14

Whole exome sequencing and data analysis. WES and analysis were performed in four members of the family (K3, K13, K14, K18). Each patient DNA was prepared according to the Illumina protocols. Briefly, 1 μg of genomic DNA was fragmented and Illumina adaptors were ligated to the fragments. Selected DNA fragments with a size of 350 to 400 bp were then PCR amplified using the TruSeq Fragment amplification kit (Illumina), and the final products were analysed for quality. The mean coverage for each sample was determined.

Table 3 | Results of the Sema4A Pro682Ser association study.

| Cohort | No. Genotypes | Frequency of allele T (%) | Two-tailed P value | OR (95% CI) |
|--------|---------------|--------------------------|-------------------|-------------|
| German and Austrian FCCTX individuals | 1,138 | 47 | 41 | 6 | 6.4 | 6.793 (2.634 to 17.518) |
| Non-cancer controls1 | 1,114 | 24 | 0 | 0.0008 | 1 |

CI, confidence interval; FCCTX, familial colorectal cancer type X; OR, odds ratio.

1Fisher’s exact test of genotype counts from cases versus controls.

The data presented here broaden our understanding of the pathophysiological role of semaphorins in human carcinogenesis and will have important consequences for screening and early tumour detection strategies of patients with FCCTX and their family members.
individuals was 23 years (range 4 to 75 years) and they all are obtained from families lacking a personal or family history of cancer. Variants were excluded if they were found in at least two individuals from the in-house databank, variants found in only one individual were further checked by functional prediction tools.

**Variant resequencing and screening of SEMA4A.** Confirmation of mutations detected at WES and screening of the SEMA4A gene in 53 further patients with FCCTX were accomplished by PCR and Sanger sequencing. Oligonucleotide primers were designed with Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) or ExonPrimer (http://ihg.helmholtz-muenchen.de/ihg/ExonPrimer. html), respectively. Primers for resequencing were designed to cover the variant and have a size preferably smaller than 300 bp. All 14 coding exons as well as intron–exon boundaries of the SEMA4A gene were analysed. Primers used in this screening are summarized in Supplementary Table 6. They were tagged by M13 sequences to facilitate direct sequencing. PCRs were performed using the HotStarTaq DNA Polymerase (Qagen) or the qeqGOLD Hot Start Mix S (PEQLAB), respectively. Capillary electrophoresis was performed on ABI PRISM 3730 DNA Analyzer or ABI PRISM 310 Genetic Analyzer, respectively (both by Applied Biosystems). Chromatograms were analysed with FinchTV v1.4.0 (Geospiza) and SeqScape software v.2.5 (Applied Biosystems).

**Reverse transcription and SEMA4A cDNA amplification.** RNA (1 μg) was digested with Dnase I, RNase-free (Thermo Scientific) and reversely transcribed with random hexamer primers using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific). A negative control (RT minus) was always included. Primers for amplification of the reference gene R2M were as previously described.28 Primers for SEMA4A transcript variants were as follows: var1-3fw, 5'-CCTG GCCCTTCTTCCTCTCC-3'; var1-4 rev, 5'-TTTCTCTGAAAGCGCAC CCC-3'; var1-4rv, 5'-TTTCTCTGACCTGTTGGCCGCC-3'. (the reverse primer was the same for all transcript variants). Primers var3-3fw and var4-4rv were also used for direct sequencing of amplified cDNA to assess mRNA expression of the V78M variant.

**Genotyping of SEMA4A Pro682Ser.** We determined the frequency of SEMA4A P682S in a normal Caucasian population and performed a genetic association analysis. Genotypes were determined by a 5′-exonuclease assay (TaqMan). Primer and probe sets were designed and manufactured using Applied Biosystems ‘Assay-by-Design’ custom service (Life Technologies. USA). General TaqMan reaction conditions were set according to the manufacturer’s instructions. Endpoint fluorescence was measured by the POLARstar plate reader (BMG Labtech). The data were exported into an Excel format and depicted and analysed as scatter plot. In this plot, genotype groups were identified as separate and distinguishable clusters. As a control for consistency of genotyping methods, determination of genotypes was repeated in at least 10% of the samples and no discrepancies were observed. Fisher’s exact test was used to test for association of genotypes from cases with genotypes from controls (GraphPad Quickcalc online; http://graphpad.com/quickcalcs/contingency1.cfm). Hardy–Weinberg equilibrium testing of cases and control was performed as previously described.48 Odds ratios were calculated using MedCalc (http://www.medcalc.org/calc/odds_ratio.php).

**Somatic cancer gene mutation screening.** Selected target regions of 50 tumour-associated genes, corresponding to 2,835 COSMIC annotated hot spot mutations, were amplified by PCR using the AmplicSeq Cancer Panel v2 (Thermo Fisher Scientific). Library preparations were performed using the Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific). Emulsion PCR and sequencing were performed with the appropriate kits (I on One Touch Template Kit v2 and Ion Proton 200 Sequencing Kit, (both from Thermo Fisher Scientific), respectively) on an Ion Torrent Proton sequencer using a single P1 semiconductor chip yielding reads ranging from 90 to 130 bp consistent with the expected PCR product size. PCR amplification was achieved by emulsion PCR and Ion Proton 200 Sequencing Kit, (both from Thermo Fisher Scientific), respectively. Capillary electrophoresis was performed on ABI PRISM 3730 DNA Analyzer or ABI PRISM 310 Genetic Analyzer, respectively (both by Applied Biosystems). Chromatograms were analysed with FinchTV v1.4.0 (Geospiza) and SeqScape software v.2.5 (Applied Biosystems).

**Vectors and transfection.** pReceiver-M46 (C-Flag + IRES-eGFP) control, SEMA4A wild-type and SEMA4A V78M mutated vectors were purchased from GeneCopoeia and propagated in One Shot TOP10 Chemically Competent E. coli (Life Technologies). Plasmids were purified by JETSTAR Maxi Plasmid Purification Kit (Genomed) and checked by direct sequencing. One day before transfection, 6 × 10^6 cells were seeded into six-well culture plates to achieve 60 to 80% confluence. Plasmid and Lipofectamine LTX (Life Technologies) were diluted in Opti-MEM medium (Life Technologies) for transfection. If not indicated otherwise, cells were usually grown for 48 h after transfection before whole-cell lysate preparation.

**Whole-cell lysates and immunoblotting.** Protein preparations were performed at 4°C. After washing cells two times with PBS, whole-cell lysates were produced from culture dish attached adherent cells using RIPA Buffer (Sigma-Aldrich) supplemented with 2 × Halt Protease Inhibitor and 2 × Halt Phosphatase Inhibitor Cocktails (Thermo Scientific) which were added just before lysis. Adherent cells were scraped from the plate after incubation in 0.25% trypsin and 0.01% EDTA for 5 to 15 min and subsequently quickfrozen in liquid nitrogen and submitted to two freeze-thaw cycles. Lysate were clarified by centrifugation at 8,000 g for 10 min. Protein concentration was determined with the DC Protein Assay (Bio-Rad) using SPECTRstar Omega and MARS Data Analysis Software (both BMG LABTECH). Lysates were diluted with 4 × Laemmli sample buffer (Bio-Rad) and 710 nM final β-mercaptoethanol and incubated for 5 min at 95°C. SDS-Polyacrylamide gel electrophoresis of equal protein amounts was performed with precast Mini-PROTEAN TGX 4-15% gels (Bio-Rad). Proteins were blotted onto low fluorescence PVDF transfer (Advanta) or supported Nitrocellulose (Bio-Rad) membranes, respectively. Membranes were blocked with 5% Non-Fat Dry Milk in TBS (Bio-Rad) with 0.01% (v/v) Tween 20 (Sigma-Aldrich). Proteins were detected with specific primary antibodies directed at: SEMA4A (1:200, #sc-67073, Santa Cruz Biotechnology), Active-β-Catenin (1:1,000, #05-665, Millipore), β-Catenin (1:200, #sc-1496, Santa Cruz Biotechnology), Akt (pan) (1:1,000, #4691, Cell Signaling), Phospho-Akt (Ser473) (1:2,000, #4060, Cell Signaling), GAPDH (1:2,000, #4370, Cell Signaling), GSK-3β (1:1,500, #9832, Cell Signaling), Phospho-GSK-3β (Ser9) (1:3,000, #9332, Cell Signaling) and anti-mouse immunoglobulins (0#260020, Dako), both diluted 1:10,000, respectively. Membranes were incubated in Restore Plus Western Blot Stripping Buffer (Thermo Scientific) at 37°C to strip antibodies. Imaging of blots was performed by chemiluminescence using WesternBright ECL horse radish peroxidase substrate (Advanta) and X-Posure films (Thermo Scientific) and CURIX 60 developer (Agfa Healthcare), respectively. ImageJ 1.47v (NIH, rsweb.nih.gov/ij) was used for analysis of band densities.
Surface expression studies. Analysis of Sema4A surface expression in ARPE-19 cells was performed as previously described11. The cDNA sequence encoding full-length mouse Sema4A (amino acids 1–760) was generated by PCR and then ligated into pEGFP-N3 (Clontech, Palo Alto, CA). Mutant Sema4A WT/EGFP construct was generated from Sema4A WT/EGFP using a QuickChange II XL site-directed mutagenesis kit (Stratagene) according to the manufacturer’s protocol. Cells were transfected using FuGENE HD (Roche).

Migration assay. A cell exclusion zone migration assay was performed with the Radius 24—Well Cell Migration Assay plate (Cell Bios) according to the manufacturer’s instructions. Briefly, HCT-116 cells were seeded into 60 mm cell culture dishes and transfected with control and Sema4A vectors. Six hours after transfection, cells from one 60 mm dish were split into four wells of one assay plate and grown overnight to allow attachment at full confluency. Time lapse microscopy was started 24 h post transfection by removing the gel spot and concurrent switching of medium to DMEM with 1% (v/v) fetal bovine serum. Cells were imaged at 24 h post transfection with the Cell Observer (Carl Zeiss) ImageJ software was used for analysis of cell migration. Closed areas were calculated for each well at different time points by subtracting the open surface area at a given time point from the open surface area at the beginning of the migration assay.

7-AAD/BrdU staining and flow cytometry. Twenty-four hours after transfection in 35 mm dishes as described, 1.5 × 10^5 HCT–116 cells were transfected to 100 mm cell culture dishes and grown for approximately 24 h under normal conditions. BrdU at a final concentration of 50 μM was then added and cells were incubated for 1 h protected from light to label actively proliferating cells. One million cells were monitored for 48 h with a 1-h interval by the Cell Observer (Carl Zeiss). ImageJ software was used for analysis of cell migration. Closed areas were calculated for each well at different time points by subtracting the open surface area at a given time point from the open surface area at the beginning of the migration assay.

Multiple sequence alignment and 3D modelling of Sema4A. Multiple sequence alignment was performed with Clustal Omega (http://www.clustal.org/omega): Structural models of Sema4A containing the Sema and PSI domains only (amino acids 55 to 527 in NP_001180229.1 reference sequence) were generated using the intensive model algorithm of phyre2 (ref. 51) and drawn by PLYVIEW-3D (http://plyview.ccmcc.hku.hk/plyview3d.html).

Statistics. Results obtained from experiments with isogenic cell lines were compared in Excel 2013 using a paired, two-tailed Student’s t-test.

References
1. Siegel, R., Desantis, C. & Jemal, A. Colorectal cancer statistics, 2014. CA-Cancer J. Clin. 64, 104–117 (2014).
2. Jasparsen, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. Gastroenterology 138, 2044–2058 (2010).
3. Lynch, H. T. et al. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin. Genet. 76, 1–18 (2009).
4. Umar, A., Risinger, J. I., Hawk, E. T. & Barrett, J. C. Testing guidelines for hereditary non-polyposis colorectal cancer. Nat. Rev. Cancer 4, 153–158 (2004).
5. Lindor, N. M. et al. Lower cancer incidence in Amsterdam-1 criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA 293, 1979–1985 (2005).
6. Steinke, V. et al. Evaluating the performance of clinical criteria for predicting mismatch repair gene mutations in Lynch syndrome: a comprehensive analysis of 3671 families. Int. J. Cancer 135, 69–77 (2014).
7. Lindor, N. M. Familial colorectal cancer type X: the other half of hereditary nonpolyposis colon cancer syndrome. Surg. Oncol. Clin. N. Am. 18, 637–645 (2009).
8. Mueller-Koch, Y. et al. Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer. Gut 54, 1733–1740 (2005).
9. Engel, C. et al. Risks of less common cancers in proven mutation carriers with Lynch syndrome. J. Clin. Oncol. 30, 4409–4415 (2012).
10. Schulz, E. et al. Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms. J. Med. Genet. 49, 422–432 (2012).
11. Sill, H., Olipitz, W., Zeisch, A., Schulz, E. & Wollter, A. Therapy-related myeloid neoplasms: pathology and clinical characteristics. Br. J. Haematol. 162, 792–805 (2011).
42. Dominguez-Valentin, M., Therkildsen, C., Da Silva, S. & Nilbert, M. Familial colorectal cancer type X: genetic profiles and phenotypic features. Mod. Pathol. doi:10.1038/modpathol.2014.49 (2014).

43. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. 30, 97–101 (2002).

44. Ruschendorf, F. & Nurnberg, P. ALOHOMORA: a tool for linkage analysis using 10K SNP array data. Bioinformatics 21, 2123–2125 (2005).

45. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595 (2010).

46. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).

47. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164 (2010).

48. Beillard, E. et al. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using ‘real-time’ quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. Leukemia 17, 2474–2486 (2003).

49. Rodriguez, S., Gaunt, T. R. & Day, J. N. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. Am. J. Epidemiol. 169, 505–514 (2009).

50. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92 (2012).

51. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363–371 (2009).

Acknowledgements
This work is dedicated to Günther J. Krejs, Professor Emeritus of Gastroenterology, Medical University of Graz, Austria. We would like to thank all patients and their family members for participating in this study. We are also grateful to Michaela Auer-Grumbach and Ian Tomlinson for critical comments, Silvia Schauer for excellent technical assistance, Wolfram Jochum for providing tumour specimens and Heinz Stammberger for supporting this project. The work was funded in part by the Austrian National Bank, Anniversary Fund (grant no. 13918), Land Steiermark, Lekämiehilfe Steiermark and ‘Vereinigung Forschungsförderung’ at Medical University of Graz, Austria (MUG).

E.S. is supported by a dissertational grant from the Austrian Society of Hematology and Oncology and the P&D programme ‘Molecular Medicine’ at MUG.

Author contributions
E.S. and H.S. designed the study. E.S., P.K., A.W. and H.S. collected family data. P.K., S.Holzapfel, S.L., J.B.G., C.R.B. and H.S. obtained patient samples and clinical data. A.R.J performed the linkage analysis. P.U. analysed the whole-exome sequencing raw data. E.S., A.L. and S. Hofer performed direct sequencing. E.H. performed and analysed dPCR. K.K. performed targeted deep sequencing and analysis. C.W. provided in-house exome data. W.R. and A.G. performed and supervised genetic association analysis. E.S. and S.N. performed in vitro experiments. E.S., W.R., A.Z., A.W., G.H., M.R.S. and H.S. interpreted results. H.S. oversaw the study. E.S. and H.S. wrote the manuscript which was reviewed and approved by all co-authors.

Additional information
Accession codes: Raw sequencing data have been deposited in the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under the accession code EGAS00001000957.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Schulz, E. et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. Nat. Commun. 5:5191 doi: 10.1038/ncomms6191 (2014).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/