Dissertation

Discovering *SEMA4A* as a novel colorectal cancer predisposition gene

Submitted by

Eduard Schulz, MD

in order to attain the academic degree

Doctor of Philosophy, PhD

in Molecular Medicine

at the

Medical University of Graz

Division of Hematology

under the supervision of

Prof. Dr. Heinz Sill

2015

DECLARATION OF ORIGINALITY

I hereby declare that this dissertation is my own original work and that I have fully acknowledged by name all of those individuals and organisations that have contributed to the research for this dissertation. Due acknowledgement has been made in the text to all other material used. Throughout this dissertation and in all related publications I followed the guidelines of “Good Scientific Practice”.

ACKNOWLEDGMENTS

This could be a very long part of my dissertation since a great many people have supported me throughout my PhD study.

I thank Heinz Sill for his continuous support and brilliant advises since my undergraduate terms.

Many thanks go also to my many helping colleagues at the Division of Hematology: Sybille Hofer, Karin Lind, Olivia Geiger, Albert Wölfler and Armin Zebisch.

I am also grateful to our cooperative partners without whom this work would not have been accomplished: Petra Klampfl, Stefanie Holzapfel, Andreas R. Janecke, Peter Ulz, Wilfried Renner, Karl Kashofer, Satoshi Nojima, Anita Leitner, Armin Gerger, Sigurd Lax, Christine Beham-Schmid, Verena Steinke, Ellen Heitzer, Jochen B. Geigl, Christian Windpassinger, Gerald Höfler, Michael R. Speicher, C. Richard Boland, Atsushi Kumanogoh.

Publication

This thesis is based on the following publication (see Appendix):

Schulz E, Klampfl P, Holzapfel S, Janecke AR, Ulz P, Renner W, Kashofer K, Nojima S, Leitner A, Zebisch A, Wölfler A, Hofer S, Gerger A, Lax S, Beham-Schmid C, Steinke V, Heitzer E, Geigl JB, Windpassinger C, Hoefler G, Speicher MR, Richard Boland C, Kumanogoh A, Sill H. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. *Nat Commun*. 2014 Oct 13;5:5191. doi: 10.1038/ncomms6191.

ABSTRACT

Background: Patients with familial colorectal cancer type X (FCRCX) fulfill clinical criteria of hereditary non-polyposis colorectal cancer (HNPCC) but lack detectable germline mutations in known cancer predisposition genes. This thesis aimed at elucidating the genetic background of this syndrome.

Methods: In a large Austrian pedigree with FCRCX, linkage analysis, whole exome sequencing and Sanger re-sequencing was used for culprit mutation identification. All coding regions of the culprit gene were then screened in 53 FCRCX patients from Austria, Germany and the USA by direct sequencing and genotype frequencies were compared with 1138 Austrian non-cancer controls genotyped by 5'-exonuclease assay. To search for co-operating mutations, targeted deep sequencing of colorectal cancers (CRCs) from culprit variant carriers was performed. HCT-116 cells were transfected with wild-type and mutant cDNAs using Lipofectamine and effects on signaling pathways were examined by immunoblotting and on proliferation by 7-AAD/BrdU flow cytometric analysis.

Results: The new p.Val78Met (c.232G>A) variant in *SEMA4A* was identified in the Austrian pedigree as a potential culprit mutation segregating with all CRCs and being associated with extracolorectal neoplasms. Screening of FCRCX patients led to the identification of two further variants (heterozygous p.Gly484Ala; homozygous p.Ser326Phe) and the SNP p.Pro682Ser in six of 48 (12.5%) German and Austrian individuals, respectively. Pro682Ser was highly associated with the FCRCX phenotype (prevalence of heterozygotes in controls 2.1% [24/1138], P=0.0009) and showed an increased risk for CRC (OR 6.79, 95% CI 2.63 to 17.52). Somatically acquired mutations in known CRC genes were identified in *TP53*, *APC*, *KRAS* and *PIK3CA*, respectively, as possible cooperating events in CRCs of *SEMA4A* V78M carriers. Compared with wild-type protein, *SEMA4A* V78M mutant demonstrated significantly higher activation of the mitogenic pathways MAPK/Erk and PI3K/Akt and increased proliferation in SEMA4A deficient HCT-116.

Conclusion: Germline variants in *SEMA4A* predispose to FCRCX.

ZUSAMMENFASSUNG

Hintergrund: Patienten mit familiärem kolorektalen Krebssyndrom Typ X (FCRCX) erfüllen die klinischen Kriterien für nicht-polypösen Darmkrebs (HNPCC), die kausalen prädisponierenden Gene sind jedoch unbekannt. Diese Doktorarbeit hat zum Ziel die genetische Ursache zu ergründen.

Methoden: Genkoppelungsanalyse, Exom-Sequenzierung (WES) und direkte Re-Sequenzierung wurden verwendet, um das vermeintliche Prädispositionsgen in einer großen österreichischen Familie zu identifizieren. Die kodierenden Regionen dieses Gens wurden anschließend in 53 Patienten mit FCRCX aus Österreich, Deutschland und den USA mittels direkter Sequenzierung untersucht und die Genotypenhäufigkeit mit 1138 österreichischen krebsfreien Kontrollprobanden mittels 5'-Exonuklease-Untersuchung verglichen. Um kooperierende Mutationen zu finden, wurde eine Tiefensequenzierung der kolorektalen Karzinomen (CRCs) der Träger aus der Familie durchgeführt. HCT-116-Zellen wurden mit Wildtyp- und mutierten cDNAs mittels Lipofectamine transfiziert, die entstehende Wirkung auf Signalwege und Proliferation durch Immunblot bzw. 7-AAD/BrdU-Durchflusszytometrie bestimmt.

Resultate: Die neue p.Val78Met (c.232G>A) Keimbahnvariante in *SEMA4A* wurde als potentiell ursächliche Mutation, die mit allen CRCs segregiert und mit extrakolorektalen Neoplasien assoziiert ist, identifiziert. Die Untersuchung der Patienten mit FCRCX führte zur Identifizierung von zwei weiteren Keimbahnvarianten (p.Gly484Ala; p.S326F) und dem Einzelbasenpolymorphismus (SNP) p.Pro682Ser in sechs von 48 (12.5%) deutschen und österreichischen Individuen. Pro682Ser war stark mit dem FCRCX-Phänotyp assoziiert (Prävalenz von heterozygoten Trägern in Kontrollen 2.1% [24/1138], P=0.0009) und zeigte ein erhöhtes Risiko für CRC (OR 6.79, 95% CI 2.63 to 17.52). Somatisch entstandene Mutationen in bekannten CRC-Genen wurden in *TP53*, *APC*, *KRAS* und *PIK3CA* als mögliche kooperierende Ereignisse in CRCs von *SEMA4A*-V78M-Trägern gefunden. Verglichen mit dem Wildtypprotein zeigte die *SEMA4A*-V78M-Mutante eine signifikant höhere Aktivierung der Signalwege MAPK/Erk und PI3K/Akt sowie eine gesteigerte Proliferation in SEMA4A armen HCT-116-Zellen.

Fazit: Keimbahnvarianten in *SEMA4A* prädisponieren zu FCRCX.

TABLE OF CONTENTS

ABBREVIATIONS

AC, Amsterdam criteria

aCGH, array-based comparative genomic hybridization

CRA, colorectal adenoma

CRC, colorectal cancer

dPCR, digital PCR

FCRCX, familial colorectal cancer type X

FFPE, formalin fixed and paraffin embedded

HNPCC, hereditary nonpolyposis colorectal cancer

JPS, juvenile polyposis syndrome

LA, linkage analysis

LOH, loss of heterozygosity

MAF, minor allele frequency

MMR, mismatch repair

MSS, microsatellite stability

MSI, microsatellite instability

NCBI, National Center for Biotechnology Information

NGS, next generation sequencing

NHLBI, National Heart, Lung and Blood Institute

ESP, exome sequencing project

PB, peripheral blood

PCR, polymerase chain reaction

PPAP, polymerase proofreading-associated polyposis

SNP, single nucleotide polymorphism

SIR, standardized incidence ratio

WES, whole exome sequencing

List of Figures

List of Tables

1 INTRODUCTION

1.1 Incidence and prevalence of colorectal cancer in Austria

The lifetime probability of developing colorectal cancer (CRC) is approximately 5% in humans (Siegel, Desantis & Jemal 2014). According to the Austrian cancer registry led by STATISTIK AUSTRIA, CRC is the third and the second most common cancer in men and women, respectively, comprising almost 4350 newly diagnosed cases in Austria in 2011. Since the start of recording in the 1980s, there is a trend for dropping incidence rates which were 26.8, 35.2 and 20.2 per 100,000 persons for all individuals, men and women, respectively, in 2011 (Figure 1a). For unknown reasons, men have a 1.7 higher cumulative risk of developing CRC and 1.8 higher mortality rate than women (Zilonke, Hackl & Baldaszti 2014). Worthwhile, also localization of CRC is sex-dependent, and involvement of the proximal colon is associated with female sex and older age hinting at distinct pathogenic mechanisms (Siegel, Desantis & Jemal 2014). Almost three out of four CRCs develop beyond 65 years of age. Roughly two thirds of cancers develop in the colon and one third in the rectum. In the ten year period between 2001 and 2011 general incidence rate dropped steadily about 25% for both sexes. Similar trends were seen in mortality rate which declined by approximately 30% and was 11.7, 15.8, and 8.7 per 100,000 persons for all patients with CRC, men and women, respectively, in 2011 (Figure 1b). Decreasing mortality eventually led to higher prevalence of CRC in Austria so that 58 % of men and 63% of women have lived for more than 6 years and approximately 40% of patients even for more than 10 years after their diagnosis until the end of 2011 (Zilonke, Hackl & Baldaszti 2014). Referred to other European countries, incidence and mortality rates are lower than the mean of all 28 members of the European Union and - with the exception of Switzerland – are the lowest of Western Europe including all geographic neighbors of Austria (Ferlay et al. 2013). Interestingly, geographic heterogeneity of incidence and mortality can also be observed within the nine different provinces of Austria. Styria, for instance, had the highest overall incidence of 33,1 per 100,000 persons but mortality was below the average of Austria (Zilonke, Hackl & Baldaszti 2014). Such geographic differences in incidence and mortality are probably more affected by environmental (diet, smoking) and social (CRC screening and treatment) than genetic factors (Lansdorp-Vogelaar et al. 2012). Indeed, the most

important factor influencing CRC survival is early detection preferably in the state of colorectal adenoma (CRA) long before any local infiltration and dissemination occurs (Siegel, Desantis & Jemal 2014, Vogelaar et al. 2006). For example, the 5-year probability of not dying from CRC after diagnosis is 70% when diagnosed at a regional stage but only 13% when diagnosed at a distant stage in non-Hispanic whites in the United States (Siegel, Desantis & Jemal 2014). In Austria, approximately 15% of CRCs had been detected on average at a distant stage during 2009-2011 and this rate has been stable since 2007 (Zilonke, Hackl & Baldaszti 2012).

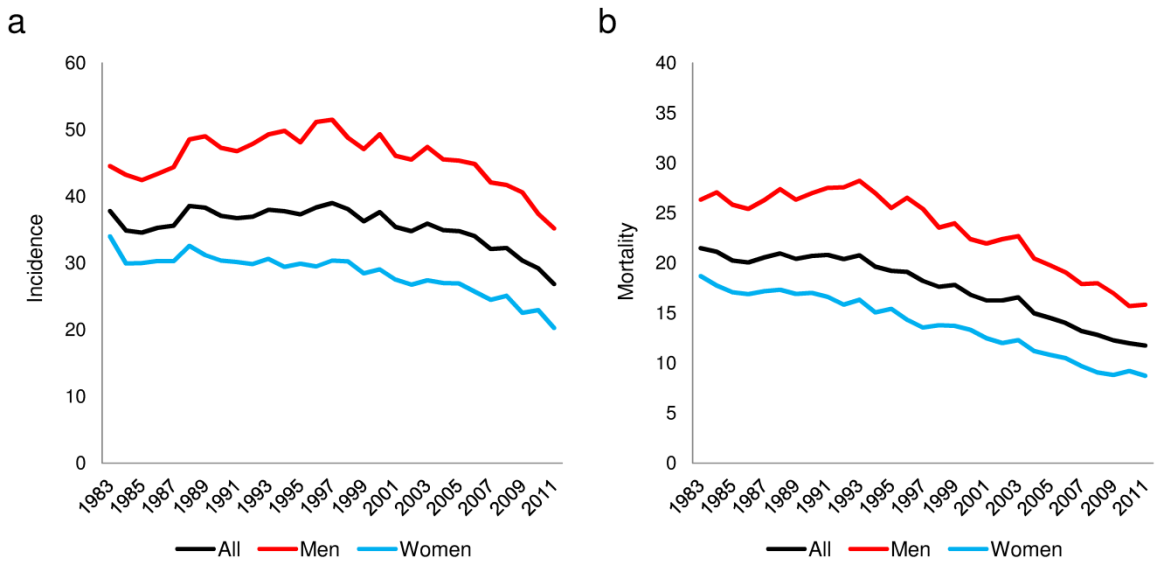


Figure 1. Incidence (a) and mortality (b) of CRC as age standardized rates of 100,000 persons, men and women, respectively. Created from data of STATISTIK AUSTRIA, Österreichisches Krebsregister, effective 25.10.2013.

1.2 Genetic predisposition to colorectal cancer

An important analysis by Lichtenstein and colleagues of almost 45,000 pairs from Scandinavian twin registries estimated that about one third of all CRCs develop due to heritable factors (Lichtenstein et al. 2000). Although this number is undoubtedly high, most CRC risk stemming from heritability can probably be attributed to a cumulative effect of multiple genes with low penetrance (Jasperson et al. 2010). Still, about 5% of all CRC cases develop as part of a Mendelian syndrome showing mostly autosomal dominant inheritance (Lynch, de la Chapelle 2003). Hereditary CRC syndromes with a high lifetime risk of CRC development and known genetic cause are summarized in Table 1. All these syndromes can be classified according to phenotypical and genetic features, and they all have in common that the onset of CRC is on average at an earlier age than in the general population (Lynch, de la Chapelle 2003). Principally and clinically based, polyposis syndromes are distinguished from the nonpolyposis syndromes. Whereas the polyposis syndromes are characterized by very early CRC development, typically before the age of 40 years, and presence of more than 100 polyps, nonpolyposis syndromes resemble more closely sporadic CRCs (Soravia et al. 1998, Hampel et al. 2005b). However, comprehensive genetic tests which are increasingly available blur this dichotomous line because there is genetic overlap between these major phenotypical manifestations. For instance, mutations in *APC*, *MUTYH* and *POLE*/ *POLD1* predisposing to (attenuated) familial adenomatous polyposis (FAP), MUTYH-associated polyposis (MAP) and polymerase proofreading-associated polyposis (PPAP), respectively, have been shown to present as nonpolyposis CRC syndromes in rare cases (Soravia et al. 1998, Morak et al. 2014, Spier et al. 2014). As this thesis originated from one particular family with predisposition to a nonpolyposis CRC phenotype, the focus is on the nonpolyposis CRC syndromes which are described in more detail.

Table 1. Validated genes predisposing to hereditary CRC syndromes with a lifetime risk of CRC greater than ten percent.

Genes	Mechanism	Syndrome	Inheritance	Lifetime CRC risk	References
Nonpolyposis					
MLH1 *MSH2* *MSH6* *PMS2* *EPCAM*[1]	DNA mismatch repair	Hereditary nonpolyposis colorectal cancer (HNPCC), Lynch Syndrome (LS) as such	AD	(15%[2]-) 35%-70%	(Hampel et al. 2005b, Stoffel et al. 2009, Senter et al. 2008, Ligtenberg et al. 2009, Ten Broeke et al. 2014)
Mixed phenotype					
POLE, POLD1	DNA repair	Polymerase proofreading-associated polyposis (PPAP), HNPCC	AD	Undetermined[3]	(Spier et al. 2014, Palles et al. 2013)
Adenomatous polyposis					
APC	Tumor suppressor	Familial adenomatous polyposis (FAP) and attenuated FAP	AD	70%-100%	(Soravia et al. 1998)
MUTYH	Base excision repair	*MUTYH*-associated polyposis (MAP)	AR	60%-90%	(Lubbe et al. 2009)
Hamartomatous or mixed polyposis					
STK11	Tumor suppressor	Peutz-Jeghers syndrome (PJS)	AD	40%-80%	(Hearle et al. 2006)
SMAD4, *BMPR1A*	Tumor suppressor	Juvenile polyposis syndrome (JPS)	AD	40%	(Brosens et al. 2007)
GREM1	Oncogene	Hereditary mixed polyposis syndrome (HMPS)	AD	Undetermined[3]	(Jaeger et al. 2012)

1) Deletion of the last exon of *EPCAM* results in multiple *EPCAM-MSH2* fusion transcripts with premature translational termination codons and methylation of the *MSH2* promoter *in cis* leading to loss of MSH2 protein and MSI.

2) The lifetime CRC risk of isolated *PMS2* mutation carriers is significantly lower compared to other MMR genes.

3) The life time risk is likely very high because mutations in respective genes have never been seen in unaffected individuals in tested families as well as very large population based screenings.

AD, autosomal dominant; AR, autosomal-recessive.

1.2.1 Hereditary nonpolyposis colorectal cancer

The first systematic approach to study individuals with increased predisposition to nonpolyposis CRC was conducted as a case study by the pathologist Alfred Warthin in 1913 (Warthin 1913). Later Henry T. Lynch *et al.* determined the mode of inheritance as autosomal dominant (Lynch et al. 1966). Eventually, the descriptive name nonpolyposis hereditary nonpolyposis colorectal cancer syndrome (HNPCC) evolved to distinguish Lynch's cancer syndrome from familial adenomatous polyposis (FAP). Since it has been already evident from the first reports that families with HNPCC have also an increased risk for extracolonic neoplasms, the term Lynch syndrome (LS) was introduced to acknowledge Henry T. Lynch's achievements in characterizing this syndrome (Boland, Troncale 1984). The underlying genetic defects of most cases of both CRC syndromes, HNPCC and FAP, were discovered during the last decade of the second millennium. Hence, hereditary CRC syndromes were more precisely divided in several different, genetically distinct cancer syndromes (Table 1). In this thesis, HNPCC is used to describe a clinically defined autosomal dominant cancer syndrome whereas LS is used to refer to HNPCC with microsatellite instability (MSI) due to germline DNA mismatch repair (MMR) gene mutations.

1.2.1.1 Genetic cause of Lynch syndrome

The DNA mismatch repair (MMR) pathway is highly conserved in eukaryotic organisms and essential for maintaining genomic stability (Li 2008). Components of this pathway serve in recognition and repair of mismatched bases resulting from DNA damage through intrinsic (e.g. random DNA replication errors, reactive metabolite byproducts such as reactive oxygen species) and extrinsic (e.g. environmental factors such as tobacco) stress factors (Li 2008). A defective MMR pathway results in hypermutation characterized by an increased rate of insertion and deletion mutations (Li 2008). This hypermutation which can be detected on a single nucleotide base level by next generation sequencing (NGS) is very specific and was recognized as a distinct genetic subtype of CRCs with deficient MMR encompassing about 15% of all CRCs (Cancer Genome Atlas Network 2012). Epithelial cells lacking a functional MMR pathway accumulate over time further cooperative oncogenic mutations, most frequently in *APC*, *TP53*, *SMAD4*, *PIK3CA* and *KRAS*, thereby undergoing accelerated

transformation from adenomas to CRCs (Cancer Genome Atlas Network 2012, Vogelstein et al. 2013). MMR deficient hypermutation becomes especially apparent in DNA sequences of simple nucleotide repeats, so called microsatellites, resulting in a genomic instability known as microsatellite instability (MSI) which can be used as a marker to identify MMR deficient CRCs (Li 2008). Testing for MSI is accomplished by PCR using five - National Cancer Institute-recommended - microsatellite markers (Umar et al. 2004a). CRCs are then defined as MSI-high (MSI-H), MSI-low (MSI-L) or microsatellite stable (MSS) if at least two, one or none of the five markers are tested positive, respectively (Umar et al. 2004a).

Disturbed MMR function and its consequence MSI is a result of either mutations or epigenetic silencing of the four major MMR genes *MLH1*, *MSH2*, *MSH6* and *PMS2* (Bronner et al. 1994, Fishel et al. 1993, Miyaki et al. 1997, Nicolaides et al. 1994). Epigenetic silencing of *MLH1* by promoter hypermethylation leading to decreased protein expression is the most common cause of MSI in sporadic CRCs and associated with an especially high rate of frameshift mutations (Cancer Genome Atlas Network 2012).

Germline mutations affecting the normal function of MMR genes are the fundamental cause of LS although transgenerational epigenetic inheritance of hypermethylated *MLH1* and *MSH2* – so called constitutional epimutation – has also rarely been described (Lynch, de la Chapelle 2003, Hitchins et al. 2011, Chan et al. 2006). Both alleles of the affected inherited MMR gene need to be inactivated to results in complete loss of MMR (Li 2008). The second hit on the wild-type allele can occur by many different mechanisms, including (large) deletion, acquired uniparental disomy, methylation or point mutation (Foulkes 2008). Loss of respective protein expression can then be detected in most cases by immunohistochemistry (IHC) techniques correlating highly with MSI (Balmana et al. 2013). The largest analysis establishing the germline mutation prevalence among CRCs – a pooled-data analysis of four large databases including 10,206 informative, unrelated CRC probands from Australia, North America, Finland and Spain - has recently shown that 3.1% (n=312) carried mutations in one of the four MMR genes (Moreira et al. 2012).

1.2.1.2 Diagnosis of HNPCC and Lynch syndrome

Diagnosis of HNPCC and differentiation from sporadic CRCs can be a challenge due to overlapping features. The International Collaborative Group on Hereditary Non-Polyposis Colon Cancer (ICG-HNPCC) defined the first criteria for the clinical diagnosis of HNPCC in Amsterdam in 1991, the so called Amsterdam I criteria (AC-I; Table 2) (Vasen et al. 1991). Using AC-I it became possible to classify patients more precisely worldwide for collaborative clinical studies. As the underlying genetic cause of HNPCC became clear in many cases during the 1990s including those cases with clear extracolonic manifestations, AC-I criteria were revised to stay abreast of changes (AC-II; Table 2) (Vasen et al. 1999). Sensitivity and specificity of these clinical criteria to identify LS have been described as high as 47% to 91% and 62% to 84% for AC-I and 77% to 81% and 46% to 68% for AC-II, respectively (Kievit et al. 2004, Win et al. 2013). It is generally accepted that families fulfilling AC-I or AC-II have HNPCC and should be genetically tested (Umar et al. 2004b).

Table 2. Clinical criteria for the diagnosis of HNPCC (Umar et al. 2004b).

Amsterdam I (1991)
Three or more relatives with CRC: • One affected patient should be a first-degree relative of the other two; • CRC found in at least two successive generations; • At least one case of CRC diagnosed before the age of 50 years; • FAP has not been diagnosed.
Amsterdam II (1999)
Three or more relatives with HNPCC-associated cancer (CRC, cancer of the endometrium, small bowel, ureter or renal pelvis) • One affected patient should be a first-degree relative of the other two; • HNPCC-associated cancer found in two or more successive generations; • At least one case of HNPCC-associated cancer diagnosed before the age of 50 years; • FAP has been excluded in any CRC cases; • Tumors should be verified by pathological examination.

As about one third of LS individuals would be missed by combined AC-I/-II and eventually excluded from genetic testing, many more guidelines were developed to identify potential LS patients with greater sensitivity (Win et al. 2013). The revised Bethesda Guidelines (Table 3) have been widely used for this purpose and are based on a cheaper MSI pretest (Umar et al. 2004a). Patients fulfilling the Revised Bethesda

Guidelines should be first tested for MSI using defined markers. Only patients with CRCs showing MSI-H should undergo genetic testing (Umar et al. 2004b). The sensitivity of the revised Bethesda Guidelines to detect the presence of *MSH2*/*MLH1* germline mutations has been 100% in a Spanish prospective, multicenter, nationwide epidemiology survey of 1,222 patients (Rodriguez-Moranta et al. 2006). Nevertheless, about 20% of LS patients would be missed if tumor analysis extended to all four MMR genes was limited to patients who fulfill Bethesda criteria (Hampel et al. 2005a, Hampel et al. 2008).

Table 3. The revised Bethesda Guidelines for testing colorectal tumors for MSI (Umar et al. 2004a).

Revised Bethesda Guidelines (2004)
Tumors from individuals should be tested for MSI in the following situations: • Colorectal cancer diagnosed in a patient who is less than 50 years of age. • Presence of synchronous, metachronous colorectal, or other HNPCC-associated tumors[1], regardless of age. • Colorectal cancer with the MSI-H histology diagnosed in a patient who is less than 60 years of age. • Colorectal cancer diagnosed in one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age 50 years. • Colorectal cancer diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.
1) Include colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, biliary tract, and brain (usually glioblastoma as seen in Turcot syndrome) tumors, sebaceous gland adenomas and keratoacanthomas in Muir–Torre syndrome, and carcinoma of the small bowel.

For this reason, newer international guidelines weaken the use of clinical criteria and recommend the use of MSI or IHC, the latter more readily available and almost equally sensitive, either in all CRCs diagnosed before the age of 70 years or in CRCs diagnosed after the age of 70 years when revised Bethesda Guidelines are met (Balmana et al. 2013, Vasen et al. 2013, Stoffel et al. 2015, Jahn et al. 2009). Since the somatically acquired oncogenic *BRAF* V600E mutation is found in sporadic CRCs with MSI due to hypermethylation of both alleles of the *MLH1* promoter, analysis of *BRAF* V600E or the *MLH1* promoter is indicated in those cases with loss of MLH1 expression or with MSI lacking IHC assessment, respectively, before germline mutation testing (Balmana et al. 2013, Vasen et al. 2013, Stoffel et al. 2015, Deng et al. 2004).

1.2.2 Features of Lynch syndrome

LS is not only characterized by increased CRC risk (Table 1) but also by increased risk of development of extracolonic cancers including carcinomas of the endometrium, ovaries, stomach, small bowel, biliary tract, urinary tract or skin (Lynch et al. 2009). Estimated lifetime risks of endometrial and ovarian cancers are 16%–54% and 1% to 24%, respectively, having equivalent clinical implications as CRC for early detection in female MMR gene mutation carriers (Bonadona et al. 2011). Cancer risks are generally higher in *MLH1* and *MSH2* compared to *MSH6* mutation carriers whereas risks are only minimally increased in *PMS2* carriers (Ten Broeke et al. 2014, Bonadona et al. 2011). Estimates of risks of less common cancers - i.e. stomach, small bowel, urinary bladder, other urothelial, breast, ovarian, and prostate cancer - were recently calculated in 2118 proven *MLH1*, *MSH2*, and *MSH6* mutation carriers from Germany and the Netherlands (Engel et al. 2012). Notably, standardized incidence ratios (SIRs) were significantly increased for all cancers but were highest for small bowel (men, SIR 250.9, 95% CI 176.7 to 345.9; women, SIR 112.2, 95% CI 65.4 to 179.7) and other urothelial cancers (men, SIR 100.4, 95% CI 65.0 to 148.2; women, SIR 121.8, 95% CI 74.4 to 188.2) compared to general population (Engel et al. 2012). Rarely, LS manifests with additional development of sebaceous adenomas and keratoacanthomas (Muir-Torre syndrome) or with brain tumors (Turcot's syndrome) (Lynch et al. 2009). Given different incidence rates of cancers in the general population, *MLH1*, *MSH2*, and *MSH6* mutation carriers have remarkable estimated 50% cumulative lifetime risks at 70 years for development of one of these cancers (Engel et al. 2012).

From a pathologic point of view, CRCs from LS patients develop more frequently in the right colon, show synchronous or metachronous growth and present as medullary, signet ring-cell or serrated carcinomas (Klarskov et al. 2012). Associated morphological features are mucin production, tumour-infiltrating lymphocytes and Crohn’s like reactions (Klarskov et al. 2012).

1.2.3 Familial colorectal cancer type X

Several studies of HNPCC patients fulfilling either AC-I or AC-II have shown that about 50% of these patients lack defects in MMR genes (Llor et al. 2005, Lindor et al. 2005, Mueller-Koch et al. 2005). Lindor *et al.* coined the term familial CRC type X (FCRCX) creating a basket for MMR proficient HNPCC of possibly different diseases (Lindor et al. 2005). Looking at FCRCX families as one group, risks for cancers other than CRC are not significantly increased contrasting with LS (Lindor et al. 2005). By definition, FCRCX is a diagnosis of exclusion and many potential genes other than MMR genes (see Table 1) have to be considered which can be accomplished nowadays in one single genetic screening using NGS.

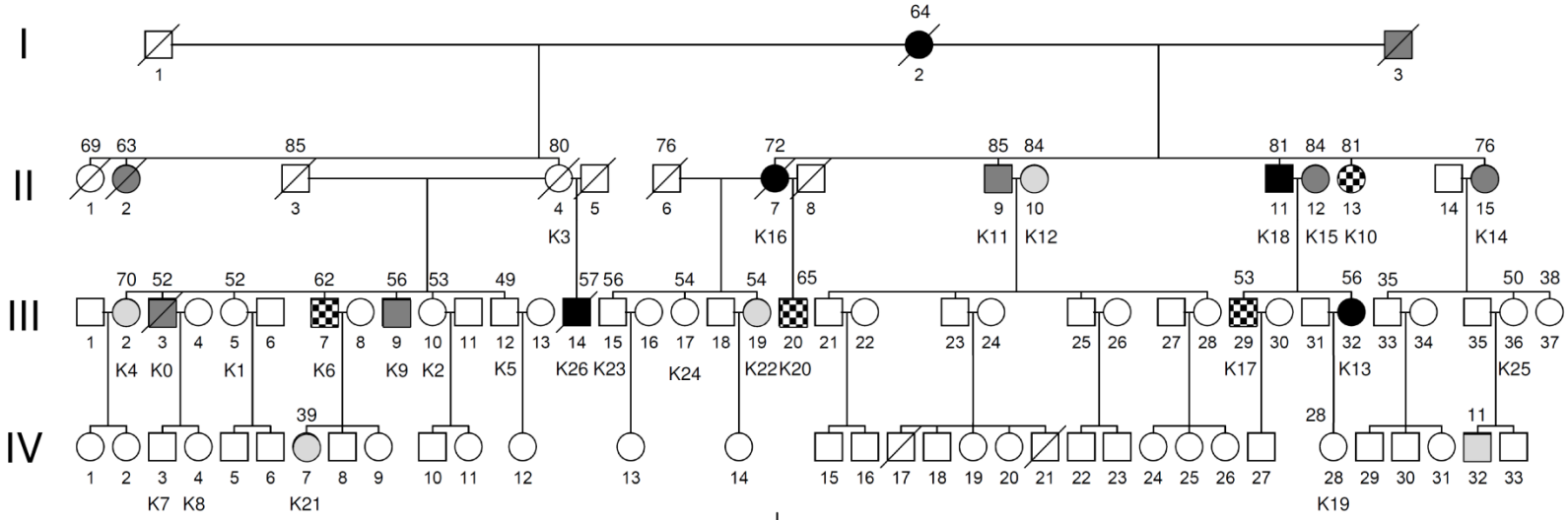
CRCs developing in FCRCX cases lack distinctive morphological features but tend to present with infiltrative growth and are associated with development in the distal colon (Klarskov et al. 2012). Histologically, they show predominantly a tubular architecture and significantly more often a conventional glandular subtype compared to LS CRCs. (Klarskov et al. 2012)

1.3 An Austrian family with colorectal cancer

During pedigree and mutational analyses searching for highly penetrant cancer predisposing mutations in patients with therapy-related myeloid neoplasms, a large Austrian pedigree meeting original AC-I was identified (Schulz et al. 2012). This family termed Family K was from southern Austria and consisted of 88 members spread into two branches (Figure 2).

The distribution of HNPCC cases in Family K pinpointed an autosomal dominant mode of inheritance with incomplete penetrance affecting three generations. In each affected individual, one to six CRAs 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 of infiltrating lymphocytes (Table 4).

As part of routine diagnostics, CRCs from two patients in either branch (K13 and K26) were tested for loss of MMR gene expression by immunohistochemical staining and for MSI by polymerase chain reaction (PCR) amplification fragment length polymorphism analysis. CRCs showed normal expression of MLH1, MSH2, MLH6 and PMS2, and were MSS (K26) or MSI-L (K13), respectively. Large genomic rearrangements in the MMR genes and *EPCAM*, respectively, as a rare cause of LS were also excluded by multiplex ligation-dependent probe amplification in patient K13. Furthermore, in individual K13 with MSI-L CRC, no somatically acquired *BRAF* V600E mutation known to be associated with somatic *MLH1* hypermethylation was found (Deng et al. 2004). Consequently, the diagnosis of FCRCX was made.



Individual	Neoplasm	Age at diagnosis (in years)
I:2	CRC	n\a
I:3	PC	n\a
II:2	Uterine cancer	31
II:7	CRA, 2 CRCs	71, 71, 72
II:9	Laryngeal cancer	79
II:10	UM	64
II:11	CRC, 6 CRAs, PC, lipoma	62, 62- 67, 74, 75
II:12	RCC	76
II:13	Angiomyolipoma, UM, CRA	53, 57, 66
II:15	Uterine tumor (unclassified), CRA, BC	46, 63, 67

Individual	Neoplasm	Age at diagnosis (in years)
III:2	UM, lipoma	46, 51
III:3	Thyroid adenoma, AML	37, 51
III:7	2 CRAs	44, 61
III:9	Testicular cancer	36
III:14	CRC	55
III:19	UM	43
III:20	CRA	55
III:29	CRA	46
III:32	UM, BC, CRC	32, 47, 48
IV:7	Breast adenoma	23
IV:32	Fetal intracardiac myoma	-

Figure 2. Extended pedigree of Family K. Black symbol, CRC; checkered symbol, CRA; dark gray, malignant neoplasm; light grey, benign neoplasm. BC, breast cancer; PC, prostate cancer; RCC, renal cell carcinoma; UM, uterine myoma; n\a, not available. Numbers above symbols indicate the age of death or the age at the time last seen (both in years, respectively).

Table 4. Clinical characteristics of colorectal neoplasms of Family K. n\a, not available.

Patient	Neoplasm	Age	Histology	Grading/Staging	Localization
K6	CRA	44	Tubular adenoma	Well to moderately differentiated	n\a
K6	CRA	61	Tubular adenoma	Well differentiated	Sigmoid colon
K10	CRA	66	Tubular adenoma	Well to moderately differentiated	n\a
K13	CRC	47	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	CRA	62	Tubular adenoma	Well to moderately differentiated	Sigmoid colon
K18	CRA	64	Tubular adenoma	Well to moderately differentiated	Ascending colon
K18	CRA	65	Tubular adenoma	Well to moderately differentiated	Descending colon
K18	CRA	66	Tubular adenoma	Well to moderately differentiated	n\a
K18	CRA	67	Tubular adenoma	Well to moderately differentiated	Descending colon
K20	CRA	55	Tubulovillous adenoma	Well differentiated	Sigmoid colon
K26	CRC	55	Adenocarcinoma	pG-2, pT-3, pN-2	Rectum

1.4 Purpose of this thesis

It was hypothesized that the FCRCX phenotype in Family K was caused by a monogenetic trait. Therefore, the present study set out to

- identify the gene and mutation causing the autosomal dominant colorectal cancer syndrome in Family K,
- investigate possible functional consequences of identified mutated gene predisposing to the inherited phenotype in Family K and
- assess the prevalence of germline variants of identified gene in different individuals with characteristics of FCRCX.

2 MATERIALS AND METHODS

2.1 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 (ethical vote 26-001 ex 13/14). 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 for molecular analyses. Patients samples used in this study included peripheral blood (PB) as well as fresh frozen and formalin fixed, paraffin embedded (FFPE) tissue specimens, some of them were provided by the Biobank of MUG.

Clinical data as well as DNA extracted from PB of 53 patients with FCRCX were provided by the German HNPCC Consortium, Bonn, Germany (n=44), the Division of Gastroenterology, Baylor University Medical Center, Dallas, Texas, USA (n=6) and the Institute of Human Genetics, MUG (n=3) in accordance with local ethical guidelines. One patient (MUG1) fulfilled modified AC whereas all others classical AC-I and/or AC-II (Umar et al. 2004b).

Caucasian Austrian control subjects (n=1138) in the genetic association analysis were recruited previously from local health screening studies (Bahadori et al. 2010). The presence of known current or previous malignant disease was excluded by patient history.

2.2 DNA and RNA isolation

DNA isolation and purification from PB mononuclear cells (MNCs), cell lines and fresh frozen tissue specimens were accomplished with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer´s instructions. The RNeasy Mini Kit (Qiagen) was used for RNA isolation from PB MNCs and cell lines. To investigate tumor-specific aberrations, tumor-bearing tissue was manually microdissected from archival, FFPE specimens and DNA isolated using the ReliaPrep FFPE gDNA Miniprep kit (Promega).

2.3 Genotyping and linkage analysis

Linkage analysis (LA) was performed in Family K afflicted with FCRCX. The GeneChip Human Mapping 250K Nsp Array (Affymetrix) was used for genotyping of family members according to the manufacturer's protocols. Briefly, 250 ng of genomic DNA was digested by NspI, ligated to adaptors by T4 DNA ligase and amplified by PCR using PCR Primer 002 and Titanium Taq Polymerase (Clonetech). PCR products of an average size between 200 and 1100 base pairs (bp) were checked by agarose gel electrophoresis, purified using the DNA Amplification Clean-Up Kit and fragmented with Fragmentation Reagent (both from Clonetech). Fragmented samples were checked for the correct size (average size below 160 bp) one more time and subsequently biotin-labeled by TdT. Labeled DNA and hybridization cocktails were mixed, denaturated and then loaded onto the array. After 16 hours of hybridization at 50°C, arrays were washed and stained by streptavidin phycoerythrin using Fluidics Station 450 (Affymetrix). The probe arrays were scanned by the GeneChip System GCS3000 (Affymetrix).

A genome-wide analysis of linkage was conducted under the assumption of an autosomal dominant mode of inheritance with assignment of phenotype to persons affected by the trait (K3, K6, K10, K13, K14, K18) and with additional inclusion of one spouse of one affected person (K15) for improved haplotype reconstruction. The disease allele was assigned a frequency of 0.001 and 100% penetrance for multipoint parametric linkage analysis on this subset of family members which was performed with the MERLIN program in the Alohomora Linkage software tool (Abecasis et al. 2002, Ruschendorf, Nurnberg 2005).

2.4 Target enrichment, HiSeq 2000 sequencing and data analysis

Whole exome sequencing (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, one 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 DNA Sample Preparation kit (Illumina), and the final products were analyzed for integrity by the Agilent Bioanalyzer. Multiple DNA libraries were combined with different indices into a single pool prior to

enrichment. Targeted regions of interest were captured by biotin-labeled probes and enriched by streptavidin beads. The enriched libraries were eluted from the beads after three washing steps. Hybridization with capture probes, washing and eluting were repeated a second time for further enrichment. Enriched targeted regions were amplified by PCR using the same primers from the TruSeq DNA Sample Preparation kit and then sequenced on a HiSeq 2000 Sequencer (Illumina).

Sequence data in FastQ format were aligned to the hg19 version of the human genome (GRCh37) using the Burrows-Wheeler Aligner (BWA; http://bio-bwa.sourceforge.net/), transformed into SAM files and then converted into compressed BAM files by picard (http://picard.sourceforge.net/) (Li, Durbin 2010). Possible PCR duplicates were marked by picard and local realignment around indels was performed using the Genome Analysis Tool Kit (GATK; http://www.broadinstitute.org/gatk/download) to prevent false positive SNPs at the end of sequence reads (DePristo et al. 2011). GATK was also used to reevaluate base quality scores, perform the raw SNP calling of all sequences within RefSeq gene exons (http://www.ncbi.nlm.nih.gov/RefSeq/) - plus ten bp at each splice site - and to recalibrate variant quality scores.

With a read length of 101 bp, there were, on average, 88333643 total reads which could be mapped to the human genome in 64.5%, respectively. The mean read depth of target regions (96.4% of RefSeq (refGene) coding exons and 97.2% of CCDS coding exons, respectively) was 49.1X. The mean coverage of target regions more than 1X was 94.5% and the mean coverage of target regions more than 10X was 87.3%, respectively.

Variant calls were annotated with ANNOVAR (http://www.openbioinformatics.org/annovar/) which contained the data from dbSNP132 (http://www.ncbi.nlm.nih.gov/SNP/) and the allele frequencies of the 1000 Genomes Project from February 2012 (http://www.1000genomes.org/) and of the ESP5400 version of the National Heart, Lung and Blood Institute (NHLBI) GO Exome Sequencing Project (https://esp.gs.washington.edu/drupal/) (Wang, Li & Hakonarson 2010). During progression of the study, variants were also manually checked for frequencies in updated versions of those databases. Furthermore, single variants were analyzed by the following prediction programs: SIFT (http://sift.jcvi.org/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/),

MutationTaster 2.0 (http://www.mutationtaster.org/), PhyloP, phastCons and GERP (the last three were precalculated from the UCSC genome browser http://genome.ucsc.edu/cgi-bin/hgGateway).
An in-house databank consisting of 18 exomes sequenced on the Illumina platform was used to exclude sequence artifacts as well as variants not covered extensively by public databases. The median age of 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.

2.5 Variant re-sequencing and screening of *SEMA4A*

Confirmation of mutations detected at WES as well as screening of the *SEMA4A* gene in the cohort of 53 further patients with FCRCX 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 re-sequencing 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 analyzed for germline muations. Primers used in this screening are provided in Table 5. They were tagged by M13 sequences to facilitate direct sequencing. Due to low complexity regions, sequencing of exon 10 could only be accomplished by a nested PCR approach. PCRs were performed using the HotStarTaq DNA Polymerase (Qiagen) or the peqGOLD 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 analyzed with FinchTV v.1.4.0 (Geospiza) and SeqScape software v.2.5 (Applied Biosystems).

Table 5. Primer sequences used in SEMA4A screening.

Oligoname	Sequence (5' to 3')	Amplicon size (bp) w\o M13
SEMA4A_Ex3_fw	tgtaaaacgacggccagAATACACACGCTTCTGCTGC	347
SEMA4A_Ex3_rv	caggaaacagctatgaccCCTCTGTTCCTCTCTCCTTCC	
SEMA4A_Ex4_fw	tgtaaaacgacggccagGGAGCACACTCAGGCAACC	347
SEMA4A_Ex4_rv	caggaaacagctatgaccCATGCACAGGCAGCCAAG	
SEMA4A_Ex5_fw	tgtaaaacgacggccagCACTAACCACCATGTCTGCTG	305
SEMA4A_Ex5_rv	caggaaacagctatgaccAATCCAGGAGATAACCGTGC	
SEMA4A_Ex6_fw	tgtaaaacgacggccagCTCTGGGGGTCCAGCAATTT	268
SEMA4A_Ex6_rv	caggaaacagctatgaccGTCCTGCATCTGGAAGGC	
SEMA4A_Ex7_fw	tgtaaaacgacggccagGATGTGAGACCTTGGCGTTC	273
SEMA4A_Ex7_rv	caggaaacagctatgaccATGCACGCATCACCACAC	
SEMA4A_Ex8_fw	tgtaaaacgacggccagAGAGGCAGGTCTGTGGAGG	281
SEMA4A_Ex8_rv	caggaaacagctatgaccGGACACATTGTACCTTCCGC	
SEMA4A_Ex9_fw	tgtaaaacgacggccagGAGGAAGCCTGTGTGTCCTG	313
SEMA4A_Ex9_rv	caggaaacagctatgaccCCACCACTCAGTCCTGCC	
SEMA4A_Ex10_ext_fw	GGCCAAACCAACGGTTTTC	1029
SEMA4A_Ex10_ext_rv	TTGGCCTTACCCTTGGCTTCT	
SEMA4A_Ex10_int_fw	TTTGTGGCAGCCATCCCTT	840
SEMA4A_Ex10_int_fw	TCCCTTTTGGAACTAGGCAGAA	
SEMA4A_Ex11_fw	tgtaaaacgacggccagAGAGAGAGCTGCTGGTGTGG	335
SEMA4A_Ex11_rv	caggaaacagctatgaccGAGCAGGCGAGATTGGTG	
SEMA4A_Ex12_fw	tgtaaaacgacggccagCAACCTGATCTGCCTCCCTC	261
SEMA4A_Ex12_rv	caggaaacagctatgaccGGCATCCTCTGCTCTAGTCC	
SEMA4A_Ex13-14_fw	tgtaaaacgacggccagTTTCTTACAGCTGGGGAGGC	574
SEMA4A_Ex13-14_rv	caggaaacagctatgaccAGCAGAAGCTGGCTGAGAAG	
SEMA4A_Ex15_fw	tgtaaaacgacggccagCCTGGCCTACCTTCTTCC	316
SEMA4A_Ex15_rv	caggaaacagctatgaccCAGATCTCAAGCCAGGCAG	
SEMA4A_Ex16-1_fw	tgtaaaacgacggccagGGCTGGGGTCCAAAGATAGG	496
SEMA4A_Ex16-1_rv	caggaaacagctatgaccATGAGGGCTCCTGAAAGCAC	
SEMA4A_Ex16-2_fw	tgtaaaacgacggccagGATCCTGAACTGGCAGGCAT	420
SEMA4A_Ex16-2_rv	caggaaacagctatgaccCATCCTAGTCAGGGCTGTGC	
M13fw	TGTAAAACGACGGCCAG	n\a
M13rv	CAGGAAACAGCTATGACC	n\a

2.6 Reverse transcription and *SEMA4A* cDNA amplification

One µg of RNA 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 *B2M* were as

previously described (Beillard et al. 2003). Primers for *SEMA4A* transcript variants were as follows: var1-3fw, 5'-CCTGGGCCTTTTCCTCTTCC-3'; var124fw, 5'-TTTCTCCTGAATGGCACCCC-3'; var1-4rv, 5'-TTTTTCTGTCACTGGCTGGC-3' (the reverse primer was the same for all transcript variants). Primers var1-3fw and var1-4rv were also used for direct sequencing of amplified cDNA to assess mRNA expression of the V78M variant.

2.7 Genotyping of *SEMA4A* Pro682Ser

Genotypes were determined by 5'-exonuclease assay (TaqMan). Primer and probe sets were designed and manufactured using Applied Biosystems 'Assay-by-Design' custom service (Life Technologies, Vienna, Austria). General TaqMan reaction conditions were according to the manufacturer of the assays. Endpoint fluorescence was measured by the POLARstar plate reader (BMG Labtech). The data were exported into Excel format and depicted and analyzed as scatter plot. In the 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.

2.8 Somatic cancer gene mutation screening

Selected target regions of 50 tumor associated genes, corresponding to 2855 COSMIC annotated hot spot mutations, were amplified by multiplexed PCR using the IonAmpliSeq Cancer Hotspot 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 (Ion 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 fragment size-range. On average, one million reads were obtained for each sample with more than 90% of bases above AQ20 and 87% to 93% reads on-target. Sequence information was obtained from tumor samples in duplicates and additionally from normal non-tumor material.

Initial data analysis was performed using the Ion Torrent Suite Software (Thermo Fisher Scientific, open source, GPL, https://github.com/iontorrent/). Briefly, this included base calling, alignment to the reference genome (hg19) using the TMAP mapper and variant calling with a modified diBayes approach taking into account the flow space information. All called variants were annotated using open source software (ANNOVAR, http://www.openbioinformatics.org/annovar/; SnpEff, http://snpeff.sourceforge.net/) and custom Perl scripts (Wang, Li & Hakonarson 2010, Cingolani et al. 2012). Coding, non-synonymous sequence variations that were detected and confirmed in tumors but not in the normal tissue were further evaluated and visually inspected in IGV (http://www.broadinstitute.org/igv/) to exclude erroneous variant calls resulting from PCR artifacts or sequence effects. Detection threshold was set to 10% mutated alleles in both duplicates.

2.9 Array comparative genomic hybridization

Tumor DNA samples were labeled using the BioPrime Array CGH Genomic Labeling System (Invitrogen) according to manufacturer's protocol. Briefly, 250 ng of AluI and RsaI digested tumor and reference DNA (Promega) were differentially labeled with dCTP-Cy5 and dCTP-Cy3, respectively (GE Healthcare) and purified by Amicon Ultra-0.5 30kDA filters (Millipore). Analysis of DNA copy number changes was conducted using a SurePrint G3 60K array (Agilent) scanned on the microarray scanner G2505B (Agilent) according to the manufacturer's instructions. Feature Extraction and DNA Workbench software (Agilent) was used for data analysis.

2.10 Digital PCR

The *SEMA4A* V78M mutation was quantitatively analyzed with digital PCR (dPCR) on the QuantStudio 3D platform (Life Technologies, USA). A Custom TaqMan SNP Genotyping Assay specific for the analysis of the V78M mutation was used and tested on a StepOne Plus instrument (Life Technologies) using the TaqMan Genotyping Master Mix (Life Technologies) according to the manufacturer's recommendations. For dPCR 17.4µl of Digital PCR Master Mix (2X) were mixed with 1.7µl of the TaqMan assay and 60ng of DNA to a final volume of 36µl and subjected to two Digital PCR 20k Chips. The chips were thermally cycled in a two-step PCR using the GeneAmp PCR

System 9700 (10min 96°C, followed by 44 cycles of 56°C 2min and 94°C 30s, final extension of 2 min 58°C) and imaged in the QuantStudio 3D instrument. Raw data were analyzed using the Relative Quantification module of the QuantStudio 3D AnalysisSuite Software. The confidence level was set to 95% and the desired precision value was 10%.

2.11 Cell culture

Adherent cell lines HT-29, SW-480, HCT-116, HRT-18 and 293T originated from ATCC were cultivated for no longer than six weeks in fully supplemented DMEM (Sigma-Aldrich) with 10% (vol/vol) HyClone FBS (Thermo Scientific) and 1X Antibiotic-Antimycotic (Life technologies) in a humidified chamber at 37 °C and 5% CO2. The identity of all cell lines was confirmed by VNTR analysis using the AmpF/STR Profiler Plus Kit and ABI PRISM 310 Genetic Analyzer (both by Applied Biosystems, respectively) according to the manufacturer's protocols, and verified at the online service of the DSMZ cell bank (http://www.dsmz.de).

2.12 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, 6x105 cells were seeded into 6-well tissue culture plate to achieve 60% to 70% confluency. On the day of transfection, plasmid and Lipofectamine LTX (Life technologies) were diluted at a ratio of 1:10 (HCT-116) or 1:5 (293T), respectively, in 500 µL serum-free Opti-MEM medium (Life technologies), incubated for one hour at room temperature and added dropwise onto the cells. For transfections in larger vessels, cell numbers and transfection reagents were scaled up according to the surface area. If not indicated otherwise, cells were usually grown for 48 hours after transfection before whole cell lysate preparation.

2.13 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 2X Halt Protease Inhibitor and 2X Halt Phosphatase Inhibitor Cocktails (Thermo Scientific) which were added just prior to lysis. Adherent cells were scraped from the plate after incubating on a shaker for 5 to 15 minutes and subsequently quickfrozen in liquid nitrogen and submitted to two freeze-thaw cycles. Lysate were clarified by centrifugation at 8000 x g for 10 minutes. Protein concentration was determined with the DC Protein Assay (Bio-Rad) using SPECTROstar Omega and MARS Data Analysis Software (both BMG LABTECH). Lysates were diluted with 4x Laemmli sample buffer (Bio-Rad) and 710 mM final β-mercaptoethanol and incubated for 5 minutes at 95°C. SDS-PAGE of equal protein amounts was performed with precast Mini-PROTEAN TGX 4-15% gels (Bio-Rad). Proteins were blotted onto low fluorescence PVDF transfer (Advansta) or Supported Nitrocellulose (Bio-Rad) membranes, respectively. Membranes were blocked with 3% (wt/vol) Non-Fat Dry Milk in TBS (Bio-Rad) with 0.01% (vol/vol) Tween 20 (Sigma-Aldrich). Proteins were detected with specific primary antibodies directed at: SEMA4A (1:200, #sc-67073, Santa Cruz Biotechnology), Active-β-Catenin (1:1000, #05-665, Millipore), β-Catenin (1:200, #sc-1496, Santa Cruz Biotechnology), Akt (pan) (1:1500, #4691, Cell Signaling), Phospho-Akt (Ser473) (1:2000, #4060, Cell Signaling), p44/42 MAPK (Erk1/2) (1:2500, #4695, Cell Signaling), Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (1:2000, #4370, Cell Signaling), GSK-3β (1:1500, #9832, Cell Signaling), Phospho-GSK-3β (Ser9) (1:3000, #5558, Cell Signaling), GAPDH (1:2000, #sc-32233, Santa Cruz Biotechnology). HRP-linked secondary antibodies were anti-rabbit IgG (#7074, Cell Signaling) and anti-mouse immunoglobulins (#P026002, Dako), both diluted 1:10000, 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 HRP substrate (Advansta), CL-XPosure films (Thermo Scientific) and CURIX 60 developer (Agfa Healtcare), respectively. ImageJ 1.47v (NIH, rsbweb.nih.gov/ij) was used for analysis of band densities.

2.14 Surface expression studies

Analysis of Sema4A surface expression in ARPE-19 cells, a human retinal pigment epithelial cell line, was performed as previously described (Nojima et al. 2013). 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 Sema4AV78M-EGFP construct was generated from Sema4AWT-EGFP using a QuikChange II XL site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. Cells were transfected using FuGENE HD (Roche).

2.15 Migration assay

A cell exclusion zone migration assay was performed with the Radius 24 - Well Cell Migration Assay plate (Cell Biolabs) according to the manufacturer's instructions. Briefly, HCT-116 cells were seeded in 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 hours post transfection by removing the gel spot and concurrent switching of medium to DMEM with 1% (vol/vol) FBS. Cells were monitored for 48h with a one hour interval by the Cell Observer (Carl Zeiss). ImageJ 1.47v 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.

2.16 7-AAD / BrdU staining and flow cytometry

Twenty-four hours after transfection in 35mm dishes as described, $1.5x10^6$ HCT-116 cells were transferred to 100 mm cell culture dishes and grown for approximately 24 hours under normal conditions. BrdU at a final concentration of 50 μM was then added and cells were incubated for one hour protected from light to label actively proliferating cells. One million cells were washed with ice cold PBS by centrifugation at 4°C and then fixed for 30 min. at room temperature, permeabilized for 10 min. on ice, re-fixed for 5 min. at room temperature, treated with DNase and finally stained

with APC anti-BrdU antibody (1:50 for 30 min. at room temperature) as well as 7-AAD according to the instructions from the APC BrdU Flow Kit (BD Pharmingen). Unlabeled native cells were used as a negative control for the APC anti-BrdU antibody. Stained cells were acquired on the BD LSR II Flow Cytometer operated with FACSDiva Software (both from BD Biosciences, respectively) with a flow rate of less than 400 cells/s on the same day of staining. Kaluza Flow Cytometry Analysis Software v1.2 (Beckman Coulter) was used for analysis and illustration of flow cytometry data.

2.17 Multiple sequence alignment and structural modeling of *SEMA4A*

Standard Protein BLAST (http://www.ncbi.nlm.nih.gov/blast/) was used to identify SEMA4A from different species in NCBI Reference Sequence Database (http://www.ncbi.nlm.nih.gov/refseq/). 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 and drawn by POLYVIEW-3D (http://polyview.cchmc.org/polyview3d.html) (Kelley, Sternberg 2009).

2.18 Statistics

Statistical comparisons between two transfected HCT-116 samples were performed in Excel 2013 using a paired, two-tailed Student's t test. 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 (Rodriguez, Gaunt & Day 2009). The odds ratios (OR) were calculated with MedCalc (http://www.medcalc.org/calc/odds_ratio.php).

3 RESULTS

3.1 Variant identification

LA of five family members with colorectal neoplasms and one unaffected, putative mutation carrier (Figure 3) was conducted to identify the genomic chromosomal location cosegregating with the FCRCX phenotype in Family K. This approach yielded four linkage regions on chromosomes 1, 3, 10 and 20 with a logarithm of odds (LOD) score of 1.5 (Figure 4), none of them harbored known cancer associated genes.

WES was then performed on four individuals from the LA (Figure 3) to identify the the specific genetic variant on the nucleotide level. 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 (Figure 5). *MUTYH* R217H was, therefore, excluded 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 (Farrington et al. 2005). No further known pathogenic or unknown mutations in known cancer susceptibility genes could be detected in any of these probands.

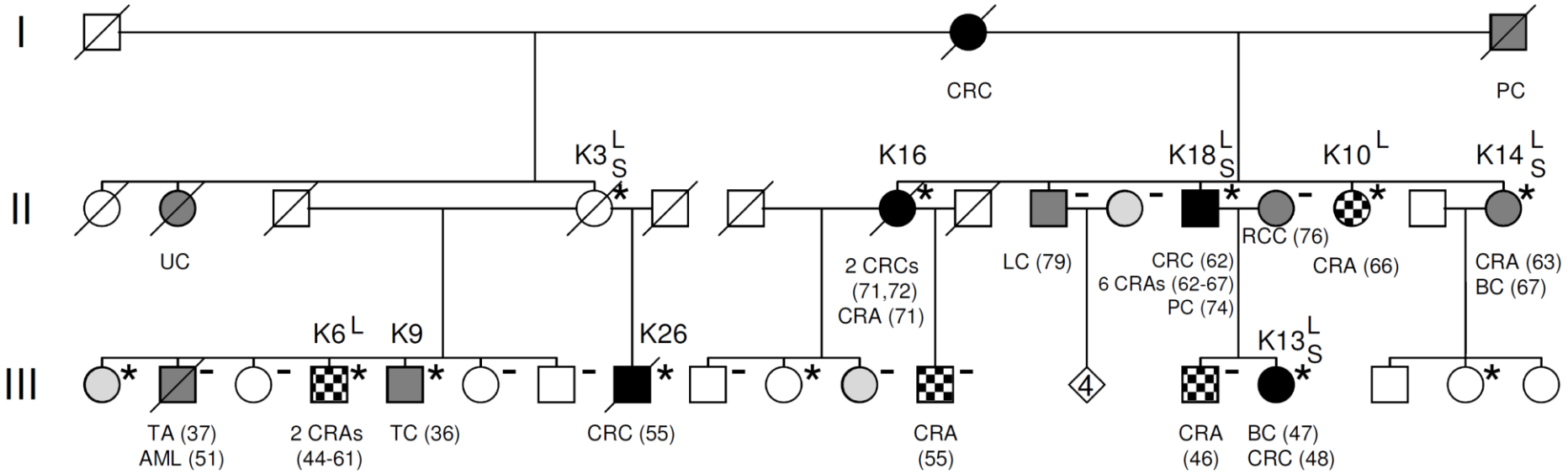


Figure 3. Familiy K with V78M mutations. L, individual included in LA; S, individual included in WES; asterisk, *SEMA4A* mutation carrier; minus, *SEMA4A* wild type; black symbol, CRC; checkered symbol, CRA; dark grey, malignant neoplasm; light grey, benign neoplasm; number in symbol, number of unspecified offspring. AML, acute myeloid leukemia; BC, breast cancer; GyT, gynaecologic tumour; OvC, ovarian cancer; PC, prostate cancer; TA, thyroid adenoma; TC, testicular cancer; UC; uterine cancer; UT, uterine tumour. Results of mutational analyses are indicated in tested individuals only. Age at diagnosis (years) is given in parentheses. For multiple colorectal adenomas, age at first presentation or at screening colonoscopy is indicated. An extended pedigree of the family is shown in Figure 2, histopathological characteristics of their colorectal neoplasms are summarized in Table 4.

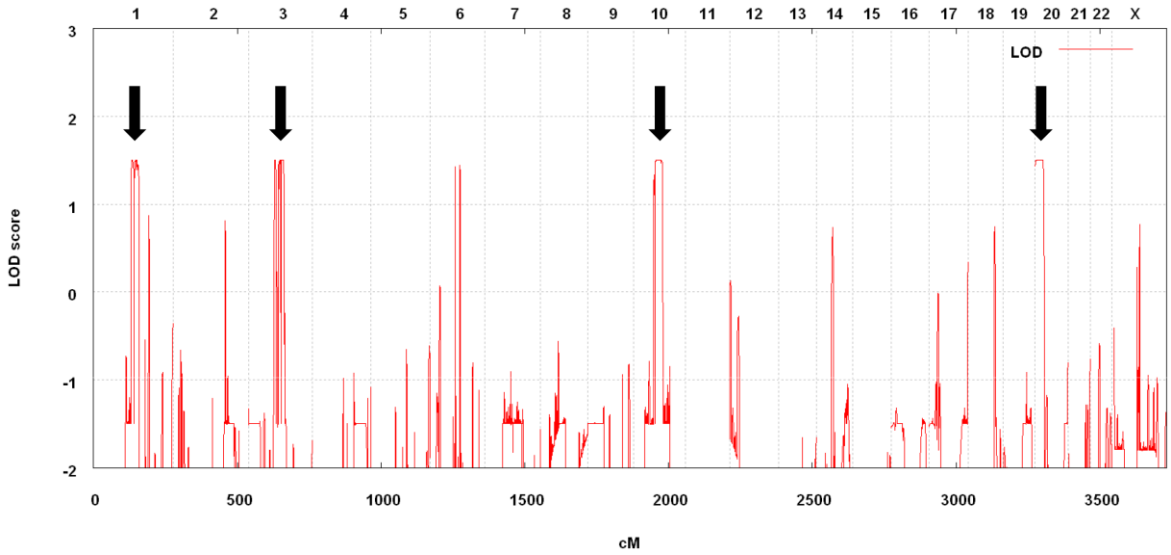


Figure 4. The panel shows parametric LOD score values in relation to the genetic position identifying four regions (arrows) of the genome reaching the maximal expected LOD score of 1.5. *SEMA4A* is located on chromosome 1q22. Human chromosomes are concatenated from p-ter (left) to qter (right) on the x-axis, and the genetic distance is given in centiMorgan (cM).

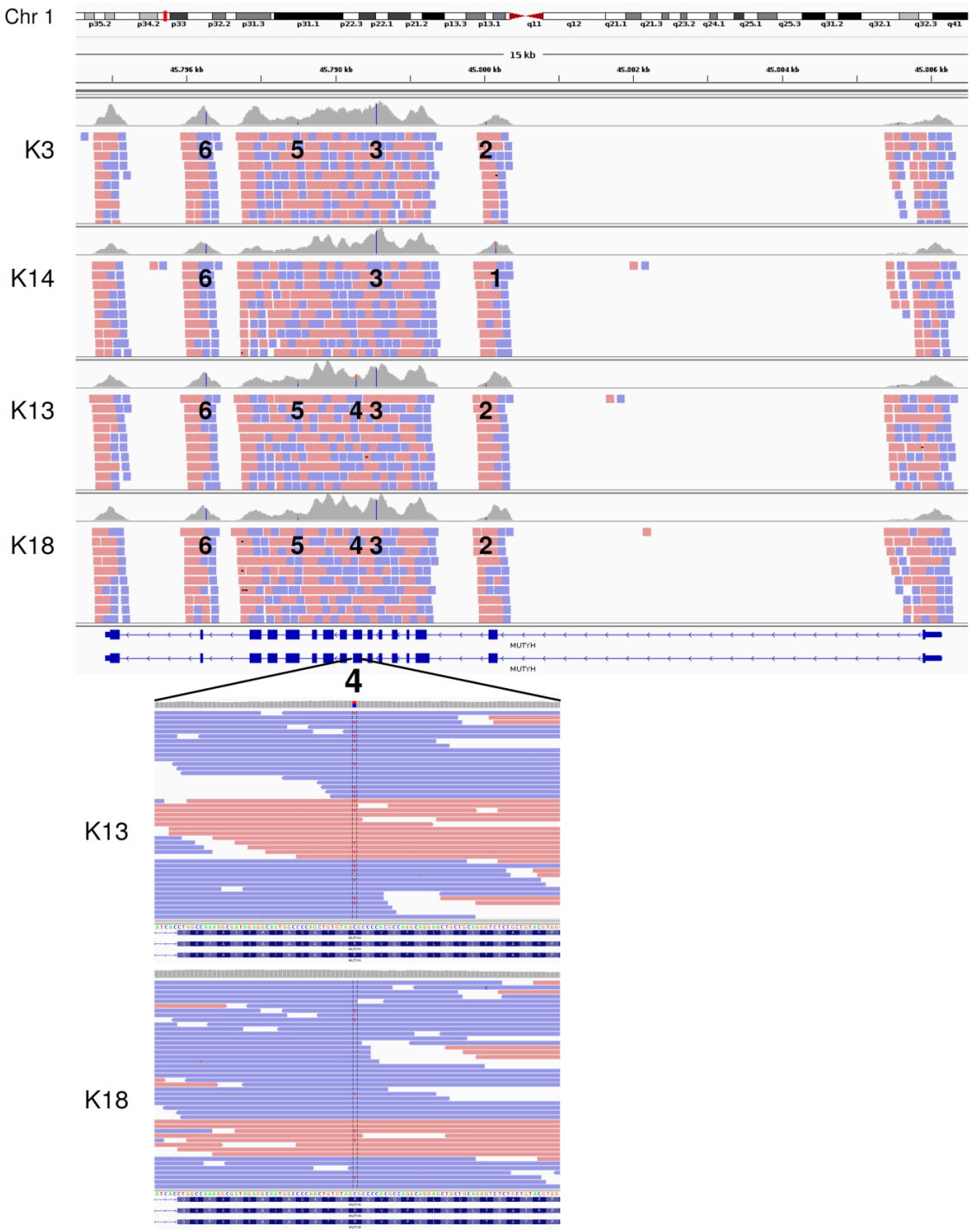


Figure 5. Analysis of the MUTYH gene in four individuals of Family K. No additional mutations other than R217H were found. The average coverage of the coding regions of *MUTYH* was 58.7x, 93.3x, 56.4x and 95.7x, and the percent coverage ≥20x was 92%, 100%, 94% and 100% in the individuals K3, K13, K14 and K18, respectively.

The R217H variation was seen in 13.25% of reads (coverage 83x) in individual K18 and 45% (60x) in K13, respectively. The sequencing data was visualized using the Integrative Genomics Viewer (IGV; Broad Institute). 1, rs3219484 (V22M); 2, rs3219485; 3, rs3219487; 4, rs147754007 (R217H); 5, rs3219489 (E335H); 6, rs3219493.

To identify novel candidate causative mutations, data of LA and WES were first combined and heterozygous, non-synonymous protein-coding or splice-site variants with a minor allele frequency (MAF) of ≤0.01 (Table 6) filtered. All variants were confirmed by Sanger sequencing and analyzed in two further family members with CRC (K16 and K26) who should share the same germline variant as part of a monogenetic trait. Only variant p.Val78Met (NM_001193300:c.232G>A) in the *SEMA4A* gene located on chromosome 1q22 was shared by all tested individuals.

In this first LA based approach, two individuals with CRA only (K6, K14) were included which constitute a frequent but not obligate part of HNPCC (Yurgelun et al. 2012). As this might constitute a potential bias, a LA independent analysis was performed focusing solely on variants from WES shared by individuals with CRC (K13, K18) or with an offspring with CRC (K3). Of 24 variants identified (Table 7) 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 (Figure 3).

Table 6. All shared heterozygous variations of individuals K3, K13, K14 and K18 from the linkage regions. Criteria: nonsynonymous SNV; exons and splice sites affected; indel, stop loss or stop gain; average heterozygosity <0.01 or unknown; not seen in in-house exome data. The chromosomal positions refer to hg19. Because of its well known function, PLAU (plasminogen activator, urokinase) was not further analyzed. n\a, not available.

Gene	ExonicFunc	AAChange	dbSNP141	Chr	Start	End	Ref	Obs
ADAMTSL4	nonsynonymous SNV	NM_019032:c.3179G>A:p.R1060H	rs147697821	chr1	150532626	150532626	G	A
CDC25B	nonsynonymous SNV	NM_004358:c.400C>T:p.R134C	rs147172963	chr20	3781128	3781128	C	T
DCBLD2	nonsynonymous SNV	NM_080927:c.1247T>C:p.V416A	n\a	chr3	98531292	98531292	A	G
GRAMD1C	nonsynonymous SNV	NM_001172105:c.472A>C:p.I158L	rs78547874	chr3	113634682	113634682	A	C
LNP1	nonsynonymous SNV	NM_001085451:c.190T>C:p.C64R	rs75122231	chr3	100170596	100170596	T	C
LNP1	nonsynonymous SNV	NM_001085451:c.194A>T:p.H65L	rs76354691	chr3	100170600	100170600	A	T
PLAU	nonsynonymous SNV	NM_001145031:c.655A>G:p.I219V	rs150389556	chr10	75673763	75673763	A	G
PLCB4	nonsynonymous SNV	NM_000933:c.1364A>G:p.K455R	rs142617224	chr20	9374275	9374275	A	G
SEMA4A	nonsynonymous SNV	NM_001193300:c.232G>A:p.V78M	n\a	chr1	156126297	156126297	G	A
SLC16A9	nonsynonymous SNV	NM_194298:c.1006G>T:p.V336L	n\a	chr10	61413778	61413778	C	A
UBOX5	nonsynonymous SNV	NM_199415:c.1427C>T:p.P476L	rs201389796	chr20	3090789	3090789	G	A

Table 7. All shared heterozygous variations of individuals K3, K13 and K18. Criteria: nonsynonymous SNV; exons and splice sites affected; indel, stop loss or stop gain; average heterozygosity <0.001 or unknown; exclusion of low quality reads; not seen in in-house exome data. The chromosomal positions refer to hg19. n\a, not available.

Gene	ExonicFunc	AAChange	dbSNP141	Chr	Start	End	Ref	Obs
ADAMTSL4	nonsynonymous SNV	NM_019032:c.3179G>A:p.R1060H	rs147697821	chr1	150532626	150532626	G	A
ANKRD28	nonsynonymous SNV	NM_001195099:c.1399A>G:p.I467V	n\a	chr3	15731538	15731538	T	C
ARSH	nonsynonymous SNV	NM_001011719:c.865T>G:p.Y289D	rs200903295	chrX	2936675	2936675	T	G
B3GALTL	nonsynonymous SNV	NM_194318:c.1325A>G:p.H442R	n\a	chr13	31898028	31898028	A	G
CDC25B	nonsynonymous SNV	NM_004358:c.400C>T:p.R134C	rs147172963	chr20	3781128	3781128	C	T
DCBLD2	nonsynonymous SNV	NM_080927:c.1247T>C:p.V416A	n\a	chr3	98531292	98531292	A	G
EPHA4	nonsynonymous SNV	NM_004438:c.2918G>A:p.R973Q	rs140023331	chr2	222290791	222290791	C	T
FILIP1	nonsynonymous SNV	NM_015687:c.1760C>T:p.S587F	n\a	chr6	76023788	76023788	G	A
FILIP1	nonsynonymous SNV	NM_015687:c.3356G>A:p.R1119Q	rs147434989	chr6	76022192	76022192	C	T
GBP5	nonsynonymous SNV	NM_001134486:c.1031C>T:p.P344L	n\a	chr1	89730487	89730487	G	A
MOGAT1	nonsynonymous SNV	NM_058165:c.655G>A:p.A219T	rs201526036	chr2	223559809	223559809	G	A
P2RY4	nonsynonymous SNV	NM_002565:c.674G>C:p.R225P	rs202027224	chrX	69478801	69478801	C	G
PODNL1	nonsynonymous SNV	NM_001146255:c.904C>T:p.R302C	n\a	chr19	14043880	14043880	G	A
PRR14L	nonsynonymous SNV	NM_173566:c.5956G>A:p.A1986T	n\a	chr22	32099580	32099580	C	T

RABGAP1L	nonsynonymous SNV	NM_014857:c.1706C>T:p.T569I	n\a	chr1	174363279	174363279	C	T
RTEL1	nonsynonymous SNV	NM_016434:c.2828C>T:p.P943L	n\a	chr20	62324333	62324333	C	T
SEMA4A	nonsynonymous SNV	NM_001193300:c.232G>A:p.V78M	n\a	chr1	156126297	156126297	G	A
SKA3	nonsynonymous SNV	NM_145061:c.1120C>T:p.L374F	n\a	chr13	21729950	21729950	G	A
SLC16A9	nonsynonymous SNV	NM_194298:c.1006G>T:p.V336L	n\a	chr10	61413778	61413778	C	A
SRSF12	nonframeshift deletion	NM_080743:c.356_358del:p.R119del	n\a	chr6	89814896	89814898	TTC	0
STT3B	nonsynonymous SNV	NM_178862:c.1150T>G:p.F384V	rs199778452	chr3	31659458	31659458	T	G
UBOX5	nonsynonymous SNV	NM_199415:c.1427C>T:p.P476L	rs201389796	chr20	3090789	3090789	G	A
ZNF660	nonsynonymous SNV	NM_173658:c.599G>A:p.G200D	rs150885666	chr3	44636284	44636284	G	A
ZNF763	nonsynonymous SNV	NM_001012753:c.634G>T:p.V212F	rs7249379	chr19	12089364	12089364	G	T

V78M lies within the SEMA domain responsible for receptor binding and Val78 is well conserved (Figure 6, Figure 7). This variant was absent from dbSNP141, the 1000 Genomes Project database and the NHLBI Exome Variant Server (ESP6500). Prediction tools favor 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). cDNA from PB leukocytes demonstrated expression of the mutant allele (Figure 8).

Phenocopy and penetrance rates of the *SEMA4A* V78M variant in Family K were estimated being 0.00 and 0.56, respectively, given a mean age of 61 years of individuals with FCRCX at disease onset (Lindor et al. 2005).

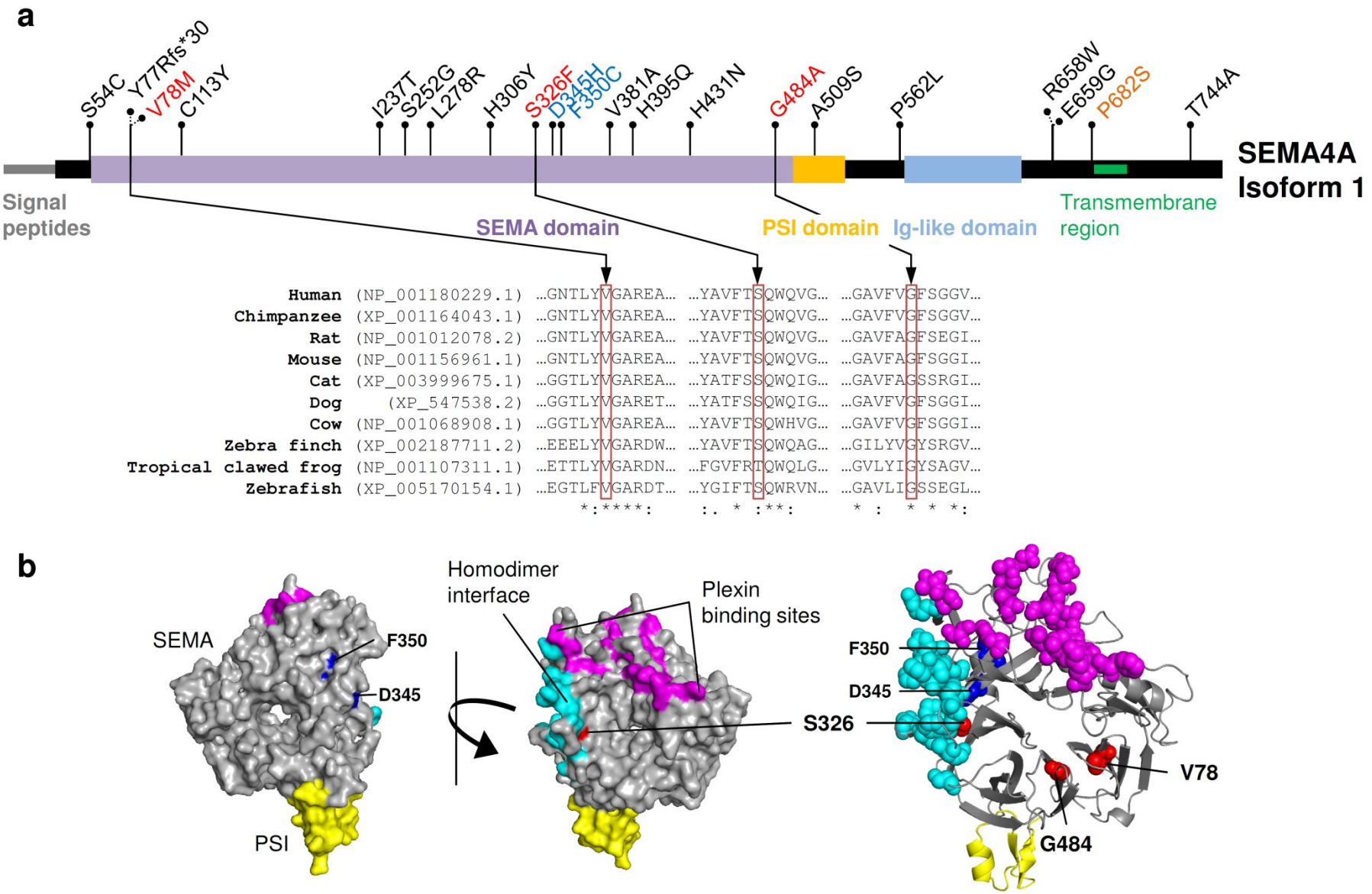

a
S54C
Y77Rfs*30
V78M
C113Y
I237T
S252G
L278R
H306Y
S326F
D345H
F350C
V381A
H395Q
H431N
G484A
A509S
P562L
R658W
E659G
P682S
T744A
SEMA4A
Isoform 1
Signal peptides
SEMA domain
PSI domain
Ig-like domain
Transmembrane region
Human (NP_001180229.1) …GNTLYVGAREA… …YAVFTSQWQVG… …GAVFVGFSGGV…
Chimpanzee (XP_001164043.1) …GNTLYVGAREA… …YAVFTSQWQVG… …GAVFVGFSGGV…
Rat (NP_001012078.2) …GNTLYVGAREA… …YAVFTSQWQVG… …GAVFAGFSEGI…
Mouse (NP_001156961.1) …GNTLYVGAREA… …YAVFTSQWQVG… …GAVFAGFSGGI…
Cat (XP_003999675.1) …GGTLYVGAREA… …YATFSSQWQIG… …GAVFAGSSRGI…
Dog (XP_547538.2) …GGTLYVGARET… …YATFSSQWQIG… …GAVFVGFSGGI…
Cow (NP_001068908.1) …GGTLYVGAREA… …YAVFTSQWHVG… …GAVFVGFSGGI…
Zebra finch (XP_002187711.2) …EEELYVGARDW… …YAVFTSQWQAG… …GILYVGYSRGV…
Tropical clawed frog (NP_001107311.1) …ETTLYVGARDN… …FGVFRTQWQLG… …GVLYIGYSAGV…
Zebrafish (XP_005170154.1) …EGTLFVGARDT… …YGIFTSQWRVN… …GAVLIGSSEGL…
b
SEMA
F350
D345
PSI
Homodimer interface
Plexin binding sites
F350
D345
S326
V78
G484

Figure 6 (page 35). 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 modeled 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.

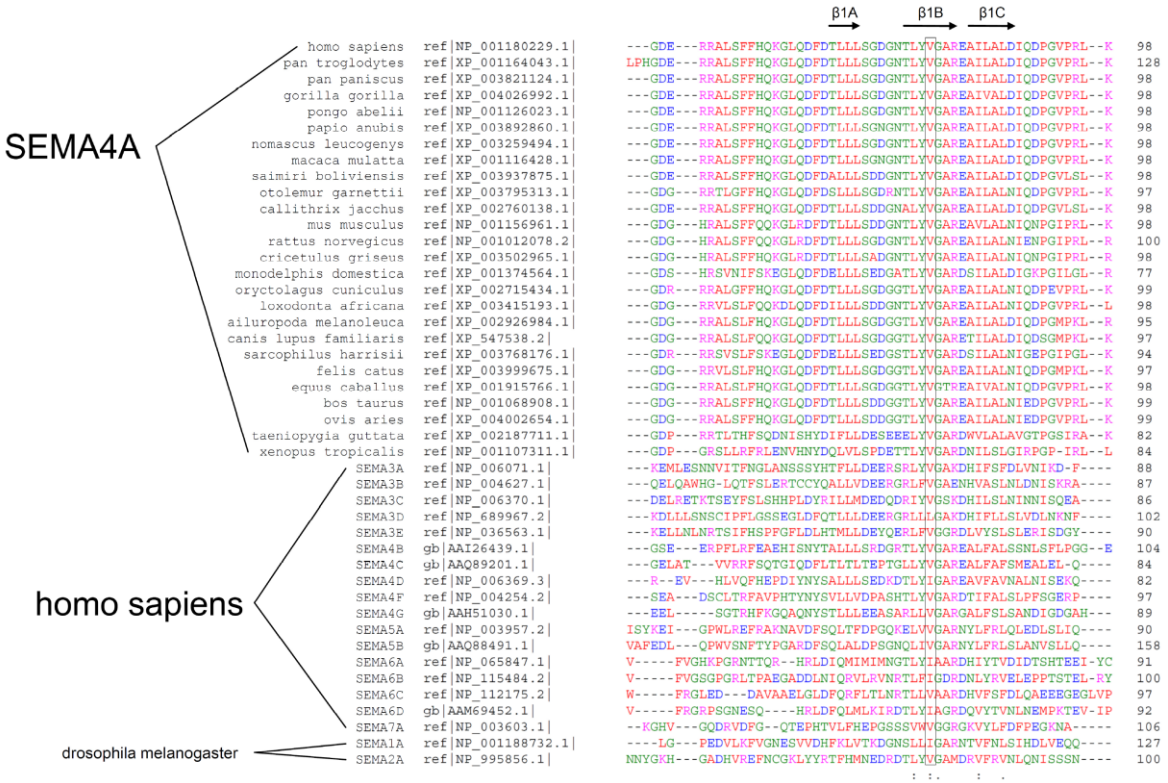


Figure 7. Multiple sequence alignment of semaphorin family members with focus on human SEMA4A Val78. Sequences of SEMA4As found in vertebrates and of all human semaphorins as well as semaphorins found in D. melanogaster were taken from NCBI's RefSeq Protein Database and aligned with Clustal Omega. Human Val78 is identical in all SEMA4As and also conserved in other classes. Secondary structure assignments derived from the crystal structures of Sema4D are adapted from (Janssen et al. 2010) **and displayed above the alignment where arrows represent β-sheets. The complete depicted sequence segment is part of the SEMA domain. Note that semaphorin classes 3 to 7 can be found in vertebrates only.**

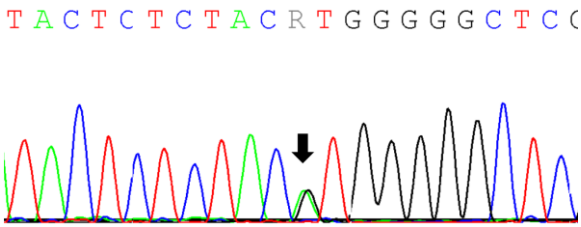


Figure 8. Sanger sequencing of amplified *SEMA4A* cDNA from lymphocytes of individual K13. The V78M mutation (c.232G>A) is normally expressed.

3.2 Genetic evaluation of CRCs from V78M carriers

To demonstrate a possible tumor suppressive nature of *SEMA4A*, CRC specimens of V78M mutation carriers were analyzed for copy number alterations (CNAs) by aCGH and loss of heterozygosity (LOH) by Sanger sequencing, respectively. Unexpectedly, gains but not deletions 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 (Figure 9). No CNAs in the *MUTYH* gene were detected in any of the three analyzed CRCs including the heterozygous R217H carrier K13.

To identify possible cooperating events, four available CRCs were also analyzed for recurrent, somatically acquired mutations in known CRC genes by targeted deep sequencing. Somatic mutations could be detected in *TP53* in 3/4, *APC* in 2/4, *KRAS* in 2/4 and *PIK3CA* in 1/4 CRC cases, respectively (Table 8). 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 (Farrington et al. 2005).

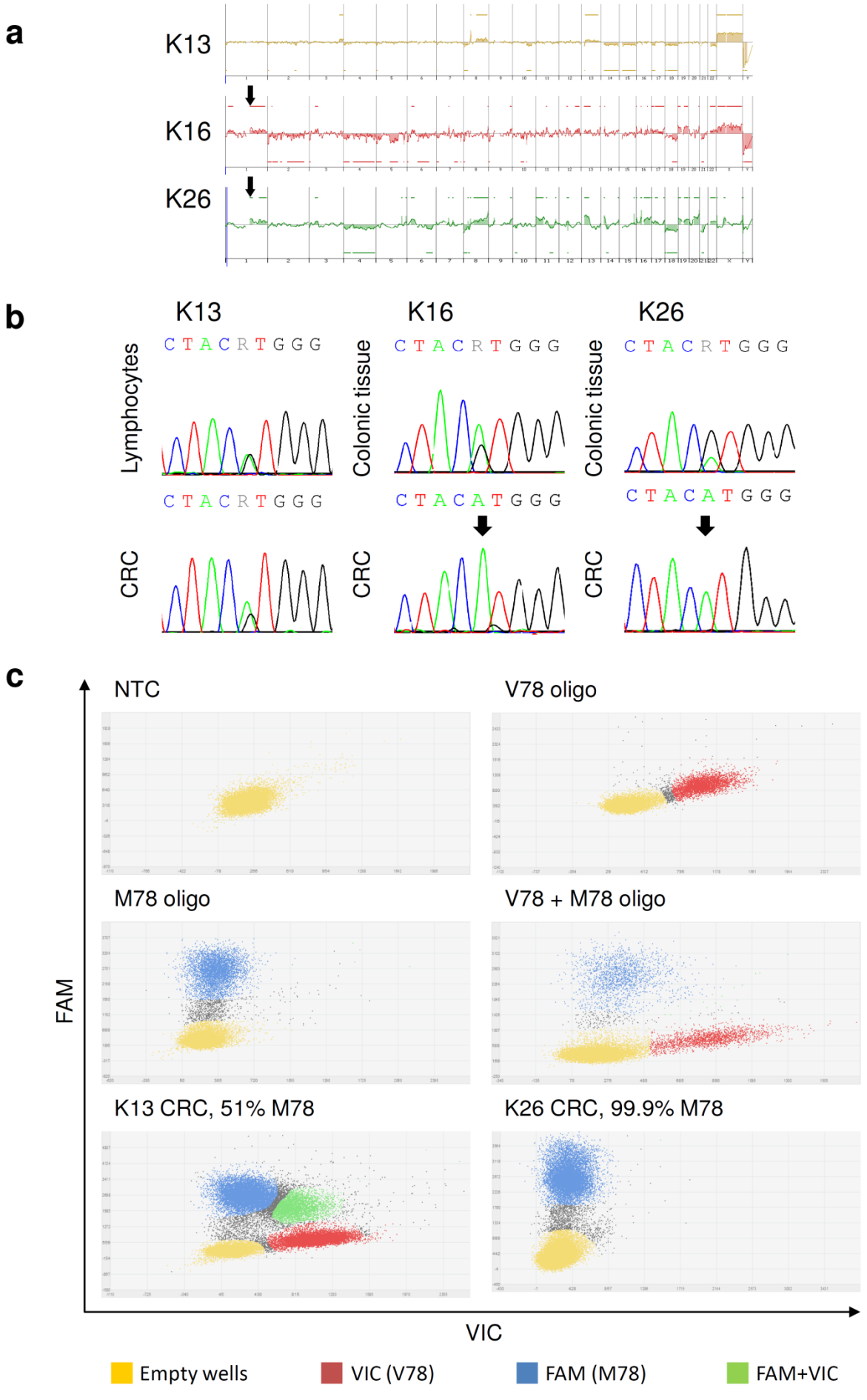

a
K13
K16
K26
b
K13
K16
K26
Lymphocytes
Colonic tissue
Colonic tissue
CRC
CRC
CRC
C T A C R T G G G
C T A C A T G G G
c
NTC
V78 oligo
M78 oligo
V78 + M78 oligo
K13 CRC, 51% M78
K26 CRC, 99.9% M78
FAM
VIC
Empty wells
VIC (V78)
FAM (M78)
FAM+VIC

Figure 9 (page 38). LOH in CRCs of patients K16 and K26. (a) aCGH 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.

Table 8. 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_000038.5:c.2626C>T	p.R876X	rs121913333
K13	APC NM_000038.5:c.4348C>T	p.R1450X	rs121913332
K13	KRAS NM_004985.4:c.34_35delinsAT	p.G12I	n\a
K13	TP53 NM_000546.5:c.380C>A	p.S127Y	n\a
K13	PIK3CA NM_006218.2:c.1633G>A	p.E545K	rs104886003
K16	None found	None found	-
K18	TP53 NM_000546.5:c.844C>T	p.R282W	n\a
K26	APC NM_000038.5:c.4135G>T	p.E1379X	rs121913326
K26	KRAS NM_004985.4:c.34G>A	p.G12S	n\a
K26	TP53 NM_000546.5:c.743G>A	p.R248Q	rs11540652

3.3 Functional assessment mutated SEMA4A^{V78M}

Family and *in vitro* as well as murine *in vivo* studies have implicated germline *SEMA4A* mutations in the development of retinal degenerative diseases (Nojima et al. 2013, Abid et al. 2006, Rice et al. 2004, Tsuruma et al. 2012). A three-dimensional protein model of human SEMA4A predicts that Val78 has no spatial relationship to residues associated with retinal disorders (Figure 6b). Mouse Sema4a is 84% identical (BLAST E-value = 0) to the human counterpart and V78 is conserved in mice. To investigate whether Sema4A^{V78M} is normally expressed on the cell surface, a fusion

gene composed of Sema4A^{V78M} and C-terminal GFP was transfected in ARPE-19 cells and the resulting GFP signal distribution inspected by fluorescence microscopy. In agreement with the computational structural prediction and the family´s history lacking apparent ocular manifestations, this fusion gene showed normal cell surface localization (Figure 10).

Figure 10. SEMA4A^{V78M} shows normal surface expression. Green, Sema4A-EGFP; red, phalloidin (actin). Representative images obtained by confocal microscopy are shown. The size of the scale bar is 20 µm.

SEMA4A is widely expressed including normal colonic tissue (Figure 11) but is undetectable on the mRNA level in 2/4 CRC cell lines analyzed (Figure 12) including HCT-116 characterized by *KRAS* and *PIK3CA* mutations. Since SEMA4A has been shown to have inhibitory effects on migration and proliferation in endothelial cells through negative modulation of the PI3K/Akt pathway, these functions were also first investigated in transiently *SEMA4A* transfected colorectal HCT-116 cells. The cell exclusion zone migration assay revealed no differences between wild-type and mutant *SEMA4A* on migration (Figure 13). However, as compared to *SEMA4A*^{wt}, significantly more *SEMA4A*^{V78M} transfected cells were in S-phase under normal growth conditions as assessed by a 7-AAD/BrdU flow cytometry assay (Figure 14).

A recent comprehensive study of mutations, copy number and mRNA expression changes in 195 CRCs showed that at least one of the three pathways PI3K/Akt, MAPK/Erk or Wnt/β-catenin is upregulated in primary, sporadic CRCs probably leading to increased proliferation and survival (Cancer Genome Atlas Network 2012). The effects of *SEMA4A* transfection on the activation of these pathways in HCT-116 cells were analyzed by western blotting of phosphorylated downstream effectors.

Compared to *SEMA4A*^{wt}, *SEMA4A*^{V78M} transfected HCT-116 cells revealed significantly enhanced activation of the PI3K/Akt and MAPK/Erk pathways (Figure 15). Transient transfection of 293T cells, however, showed no effect of SEMA4A on the PI3K/Akt pathway (Figure 16).

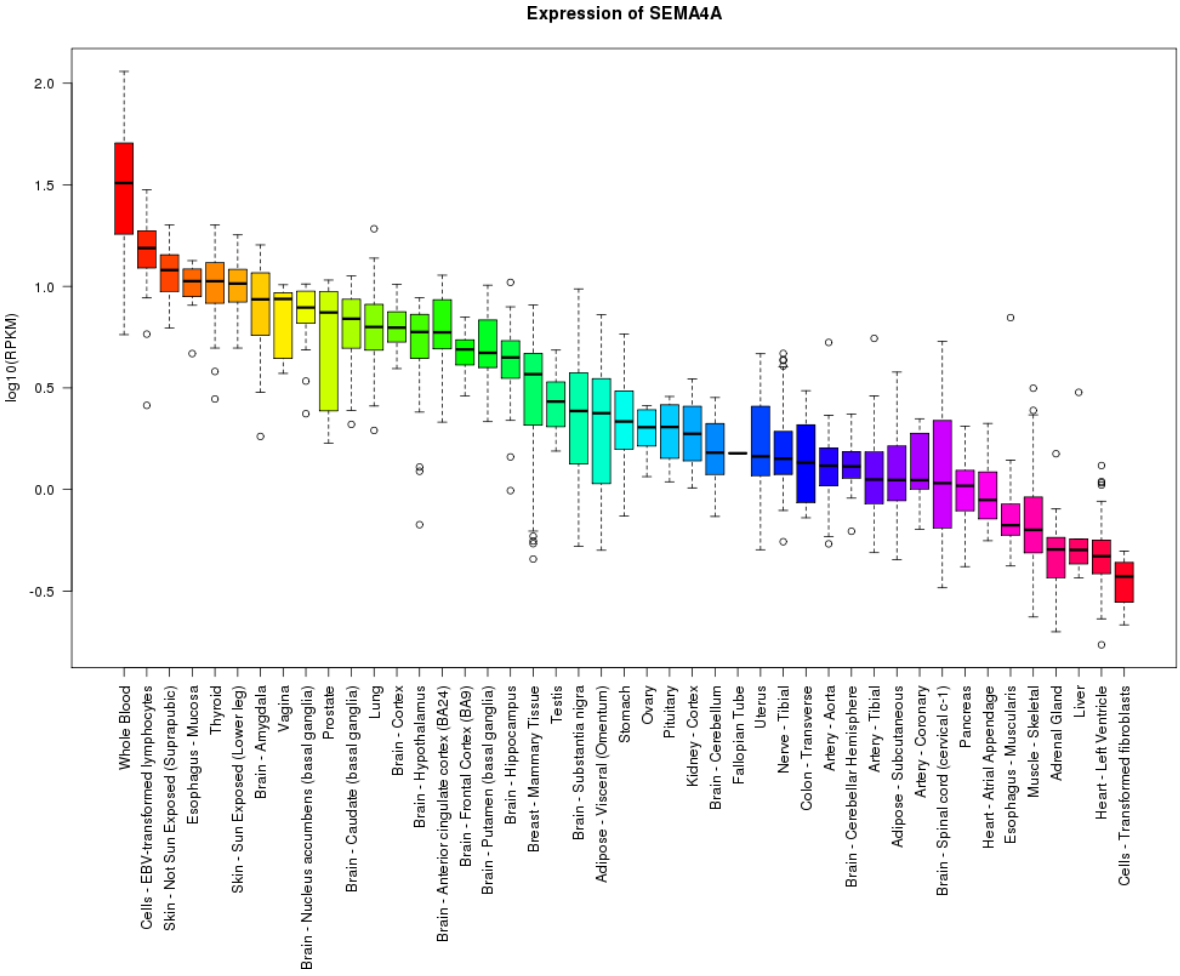


Figure 11. mRNA expression of *SEMA4A* in different human tissues. Data was extracted from the Genotype-Tissue Expression Portal (GTEx; http://www.broadinstitute.org/gtex/).

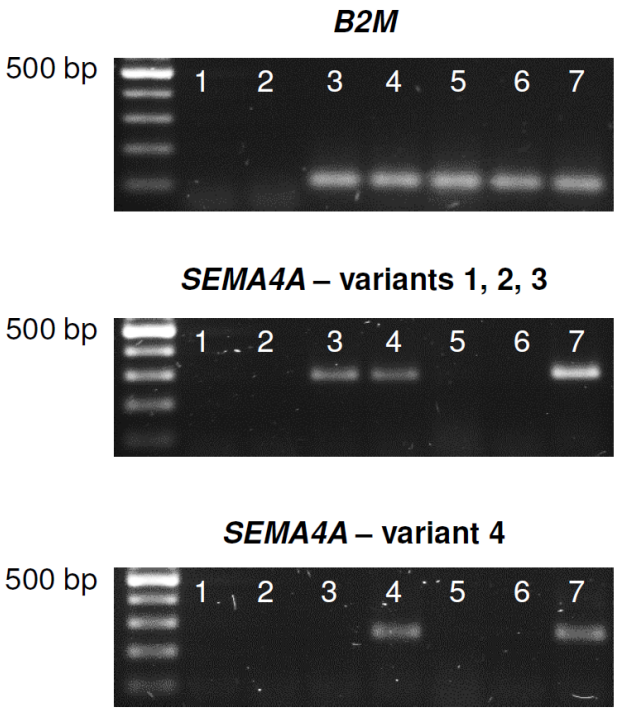


Figure 12. Expression of *SEMA4A* in CRC cell lines. Transcript variants of both *SEMA4A* isoforms were amplified. *B2M* was used as a reference gene. Jurkat cells (acute T-ALL) were used as a positive control. 1, Blank; 2, reverse-transcriptase negative control (Jurkat mRNA); 3, Jurkat; 4, HT-29; 5, SW-480 ; 6, HCT-116; 7, HRT-18.

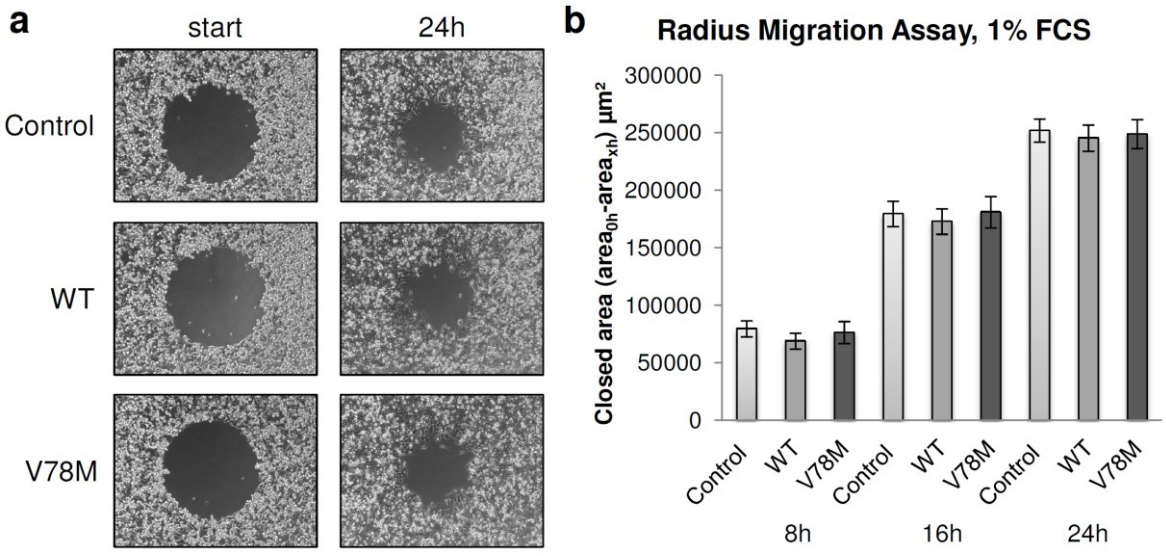


Figure 13. Radius migration assay of *SEMA4A* transfected HCT-116 cells. (a) Representative microscopic pictures from one experiment. (b) Results of three experiments performed as quadruplicates. Bars represent means ± s.e.m.

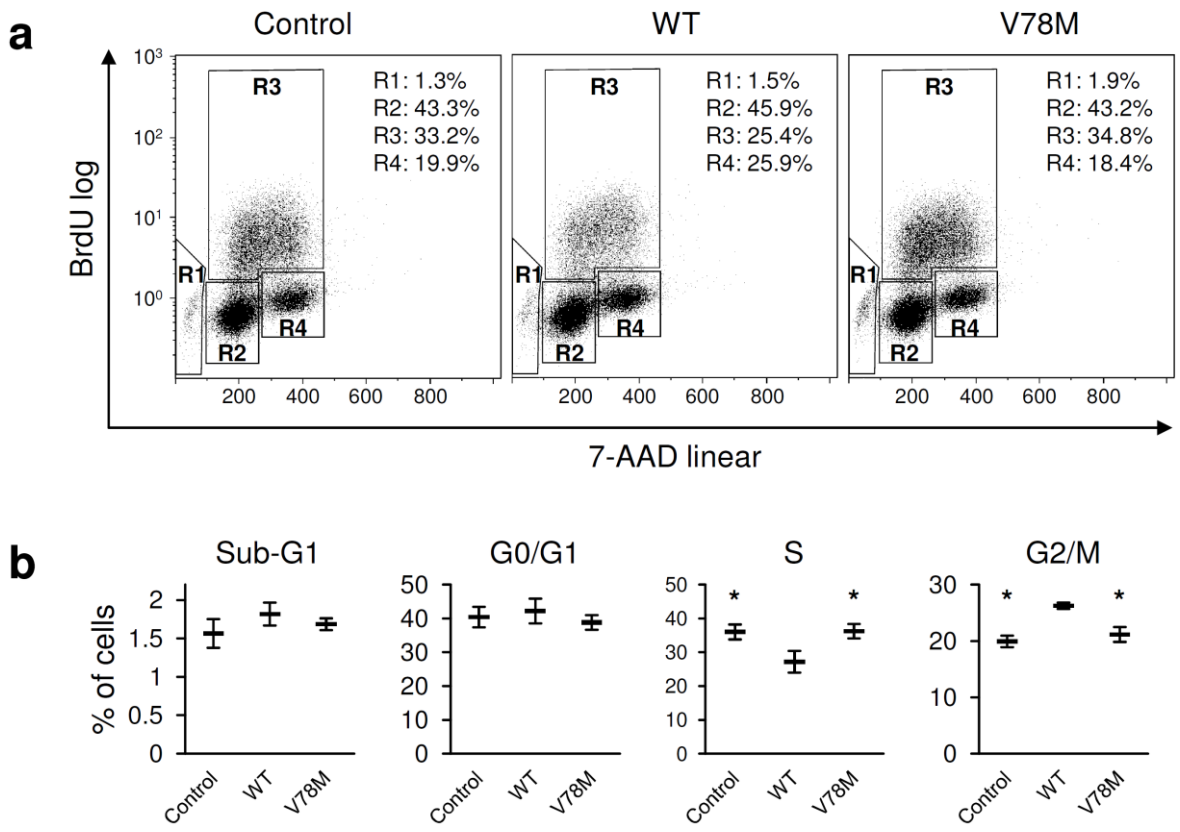


Figure 14. SEMA4A^{V78M} leads to cell cycle changes in HCT-116 cells. Representative density plots and statistical analysis of GFP positive *SEMA4A* transfected HCT-116 cells stained by 7-AAD and APC-anti-BrdU antibodies for cell cycle analysis. Cells were analyzed 48h after transfection. Significantly more *SEMA4A*^{V78M} than *SEMA4A*^{WT} 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 to WT). Cell cycle phase: Sub-G1 (R1), G1/G0 (R2), S (R3), G2/M (R4).

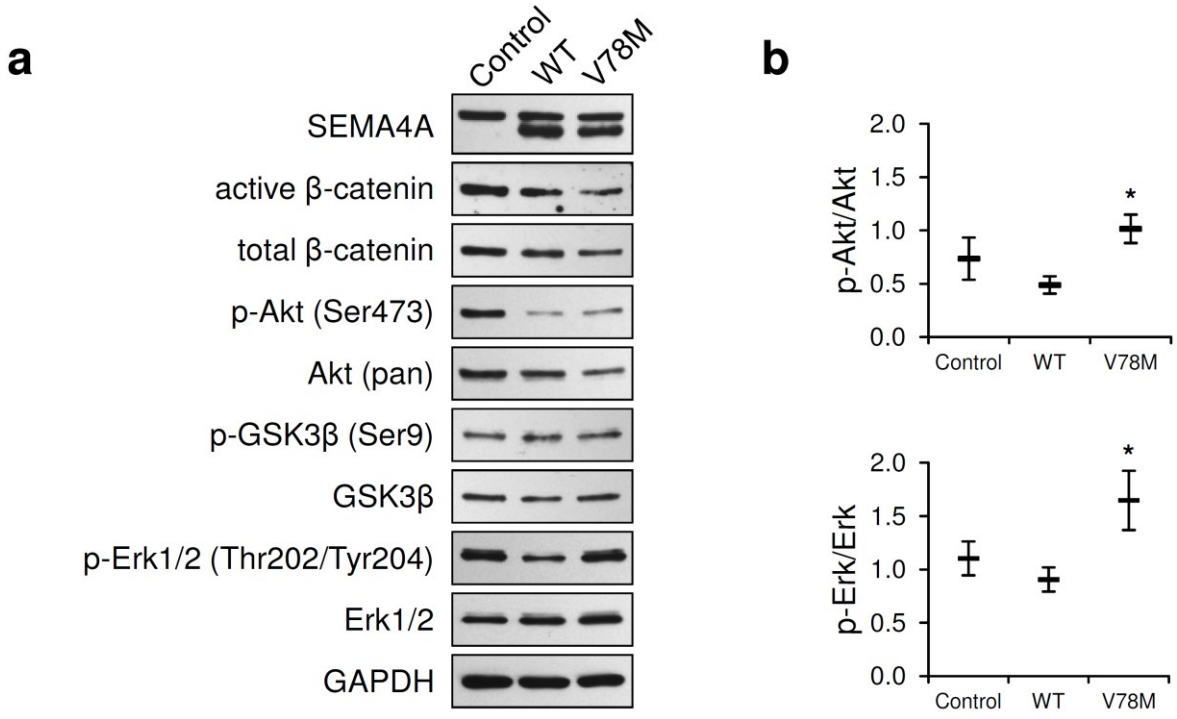


Figure 15. SEMA4A^{V78M} leads to increased activation of PI3K/Akt and MAPK/Erk pathways in HCT-116 cells. Representative immunoblots and statistical analysis of *SEMA4A* transfected HCT-116 cells (whole cell lysates) lyzed 48h after transfection. *SEMA4A*^{V78M} transfected cells show increased phosphorylation of Akt and Erk (mean ± s.e.m.; n =6 per group; two-tailed paired Student's t test; *, p <0.05 compared to 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.

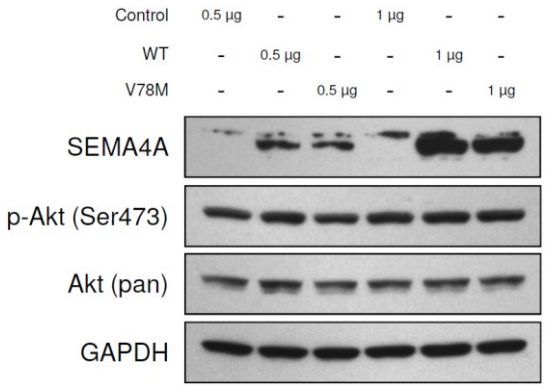


Figure 16. Immunoblots of *SEMA4A* transfected 293T cells showing a representative result of two independent experiments. No effects on PI3K/Akt pathway were seen after transfection of *SEMA4A*. The MAPK/Erk pathway is not constitutively active in this cell line.

3.4 Mutational screening of *SEMA4A*

To study the prevalence of *SEMA4A* germline mutations in FCRCX, 53 unrelated FCRCX cases from Austria, Germany and the US (Table 9) were screened, and two further mutations located in the SEMA domain (homozygous c.977C>T, p.Ser326Phe; heterozygous c.1451G>C, p.Gly484Ala, rs148744804) were identified (Figure 17). These mutations affect highly conserved residues (Figure 18, Figure 19) and prediction tools indicate an effect on protein function for both of them (Table 10).

The G484A variant found in index patient BN04 who had been analyzed without success for germline mutations in *MLH1*, *MSH2*, *MLH6* and *PMS2* by direct sequencing has a global MAF of 0.001 in the 1000 Genomes Project and ESP6500 databases. It was also found in her brother affected with CRC showing MSI-L (Figure 17**,** Figure 20).

The novel S326F variant found in index patient BN01 with MSS CRC affects a residue predicted to be involved in homodimer formation (Figure 6**,** Figure 19), a process which is required for target receptor binding (Cagnoni, Tamagnone 2013).

In addition, mutational screening of *SEMA4A* revealed an increase in the prevalence of Pro682Ser (c.2044C>T, rs76381440) which was 13% (6/47) in Austrian and German FCRCX patients (Table 9**,** Figure 17). A subsequent genetic association study using DNA from 1138 Caucasian control subjects from Austria without a personal or family history of cancer demonstrated that P682S had a highly significant association with the FCRCX phenotype (Table 11). Screening the 1000 Genomes Project database revealed a comparable prevalence of heterozygotes among European individuals of 2.0% as compared to the Austrian control population.

Table 9. Patients' characteristics and *SEMA4A* screening results. Germline mutations and CRC associated SNP are in bold. Variant coding refers to NG_027683.1, NM_001193300.1 and NP_001180229.1, respectively. Age is depicted in years. Ams, Amsterdam criteria group; CRA, colorectal adenoma; DCIS, ductal carcinoma in situ; EnC, endometrial cancer; PaC, pancreatic cancer; OvC, overian cancer.

UPN	Sex	Tumor	Age	Ams	SEMA4A variants
MUG1	M	CRC	28	Modified	g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
MUG2	M	PaC	47	I+II	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
MUG3	M	CRC	13	I+II	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN01	F	CRC	43	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573); g.16569C>T, c.977C>T, p.Ser326Phe, hom.
BN02	F	CRC	30	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN03	M	CRC	47	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN04	F	OvC	27	I	g.30159G>C, c.1451G>C, p.Gly484Ala, het. (rs148744804)
BN05	F	CRC	40	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN06	F	CRC	64	I	-
BN07	F	CRA	45	I	-
BN08	M	CRC	53	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN09	M	CRC	53	I	g.31955G>A, c.2187G>A, p.Pro729=, het. (rs41265019)
BN10	F	CRC	75	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN11	M	CRC	61	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573); g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
BN12	F	CRC	76	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN13	F	CRC	46	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN14	M	CRC	49	I	-
BN15	F	CRC	68	I	-
BN16	F	CRC	82	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN17	F	CRA	64	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)

BN18	M	CRC	36	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN19	F	CRC	42	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN20	F	CRC	40	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN21	M	CRC	52	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN22	M	CRC	45	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573); g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
BN23	F	CRC	45	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN24	F	CRC	49	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN25	F	CRC	43	I	-
BN26	M	CRC	48	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573); g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
BN27	F	CRC	54	I	-
BN28	F	CRC	66	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN29	F	EnC	62	II	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN30	M	CRC	45	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN31	F	CRC	51	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN32	F	DCIS, CRC	38, 40	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573); g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
BN33	M	CRA	48	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN34	F	CRC	40	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN35	M	CRC	86	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN36	M	CRC	47	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN37	F	CRC	56	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN38	M	CRC	49	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BN39	F	CRC	50	II	-
BN40	M	CRC	60	I	g.31906G>A, c.2138G>A, p.Arg713Gln, het. (rs41265017)
BN41	M	CRC	48	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)

BN42	M	CRC	51	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573); g.31812C>T, c.2044C>T, p.Pro682Ser, het. (rs76381440)
BN43	M	CRC	42	II	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BN44	M	CRC	44	II	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BC17	F	CRC	56	I+II	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BC20	F	CRC	47	II	-
BC41	F	CRC	47	I	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BC95	F	CRC	47	I+II	g.31484C>T, c.1716C>T, p.Pro572=, hom. (rs12401573)
BC663	M	CRC	47	I	g.31484C>T, c.1716C>T, p.Pro572=, het. (rs12401573)
BC676	M	CRC	59	I	-

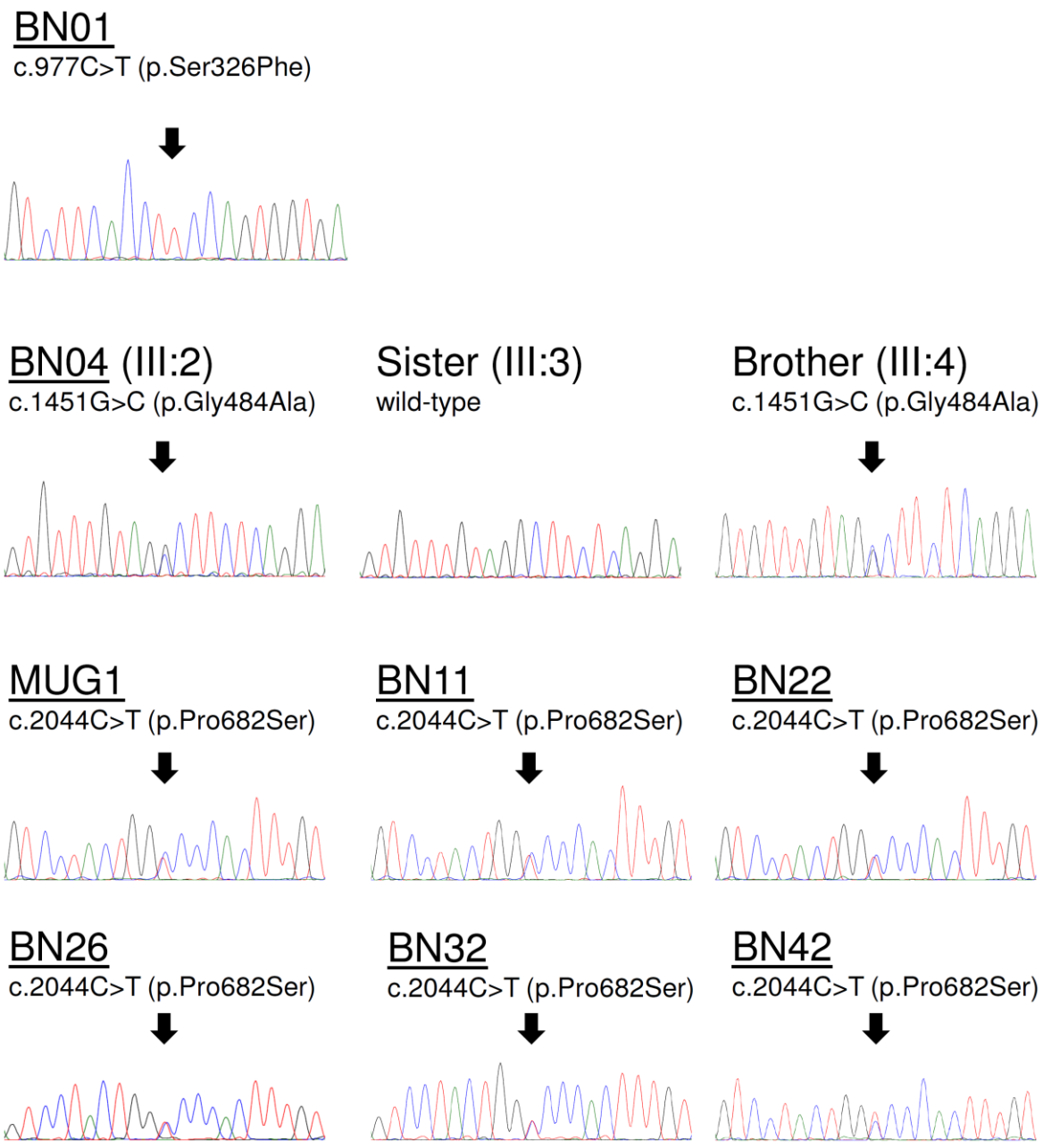


Figure 17. Further *SEMA4A* germline variants identified in a screening of FCRCX patients.

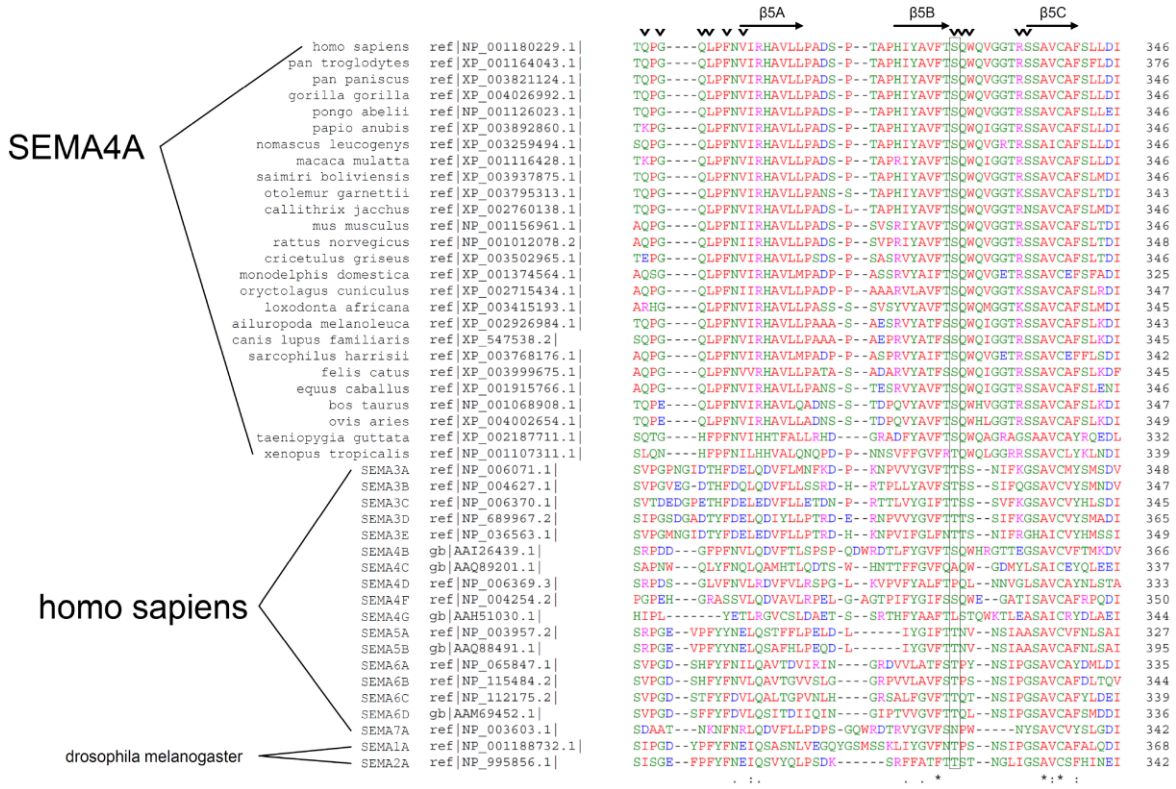


Figure 18. Multiple sequence alignment of semaphorin family members with focus on human SEMA4A Ser326. Sequences of SEMA4As found in vertebrates and of all human semaphorins as well as semaphorins found in D. melanogaster were taken from NCBI's RefSeq Protein Database and aligned with Clustal Omega. Secondary structures assignments derived from the crystal structures of Sema4D are adapted from Janssen et. al. and displayed above the alignment (Janssen et al. 2010). Arrows represent β-sheets and arrowheads indicate residues involved in homodimer formation (adapted from NCBI's RefSeq Protein Database). The complete depicted sequence segment is part of the SEMA domain. Note that semaphorin classes 3 to 7 can be found in vertebrates only.

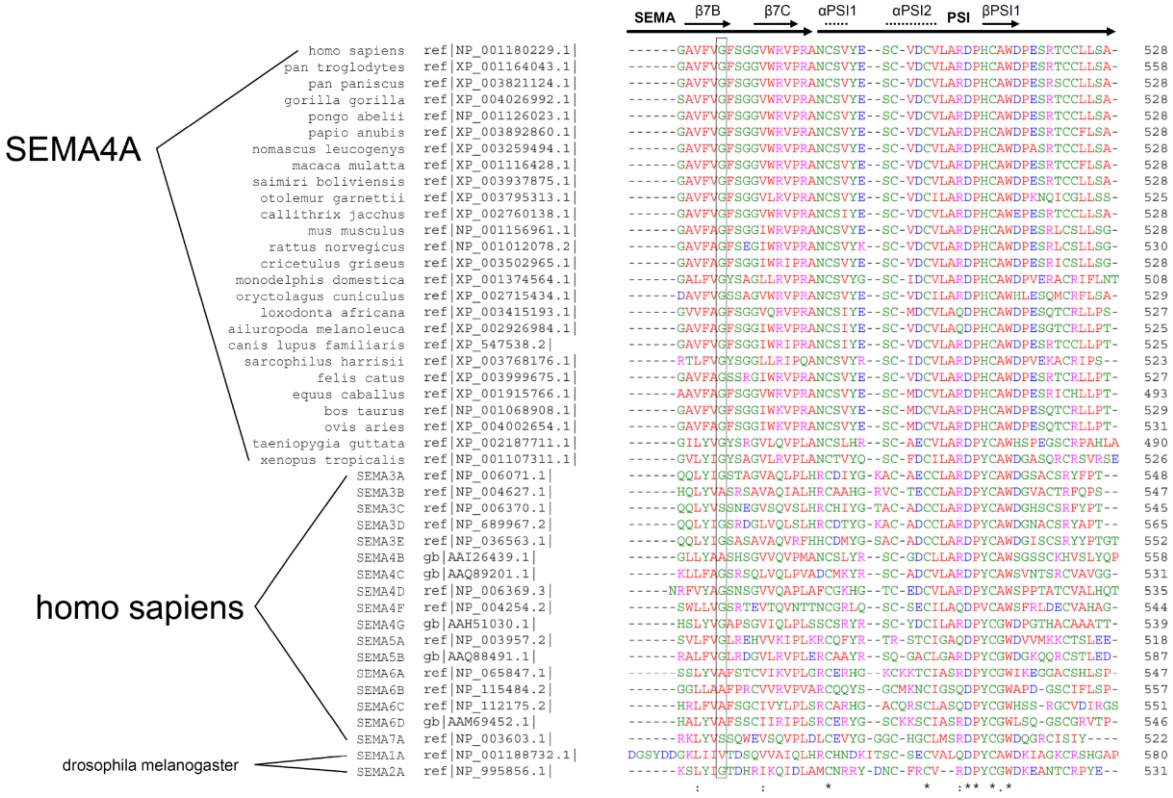


Figure 19. Multiple sequence alignment of semaphorin family members with focus on human SEMA4A Gly484. Sequences of SEMA4As found in vertebrates and of all human semaphorins as well as semaphorins found in D. melanogaster were taken from NCBI's RefSeq Protein Database and aligned with Clustal Omega. Human Gly484 is identical in known SEMA4As. Secondary structures assignments derived from the crystal structures of Sema4D are adapted from Janssen et. al. and displayed above the alignment (Janssen et al. 2010). Small arrows represent β-sheets and dotted lines α-helices. Note that semaphorin classes 3 to 7 can be found in vertebrates only.

Table 10. Mutation predictions of germline *SEMA4A* mutations.

mRNA	Protein	Mutation Taster 2 probability value	Mutation Taster 2	PP2 HumDiv score	PP2 HumDiv	PP2 HumVar score	PP2 HumVar	SIFT Score	SIFT	phyloP 100way	phastCons 100way	phastCons Elements 100way	GERP (position -specific score)
c.232G>A	p.Val78Met	0.95	Disease causing	1	Probably damaging	0.987	Probably damaging	0	Damaging	7.434	1	lod=65, 407	2.94
c.977C>T	p.Ser326Phe	1	Disease causing	0.995	Probably damaging	0.945	Probably damaging	0.01	Damaging	5.577	1	lod=27, 320	4.45
c.1451G>C	p.Gly484Ala	1	Disease causing	0.876	Possibly damaging	0.541	Possibly damaging	0.19	Tolerated	4.124	1	lod=59, 397	4.27

Table 11. Results of the *SEMA4A* Pro682Ser association study.

Cohort[1]	No.	CC	CT	TT	Frequency of allele T	Two-tailed P value[3]	OR (95% CI)
German and Austrian FCRCX individuals	47	41	6	0	6.4%	0.0008	6.793 (2.634 to 17.518)
Non-cancer controls[2]	1138	1114	24	0	1.1%	-	1

1) Genotypes of cases and controls were in Hardy-Weinberg equilibrium, with p values (df=1) of 0.718 and 0.647, respectively.
2) Men, n=574 (50.4%); women, n=564 (49.6%); mean age: 60 years (+/- s.d. 18); median age: 64 years (range 15-99).
3) Fisher's exact test of genotype counts from cases vs. controls.

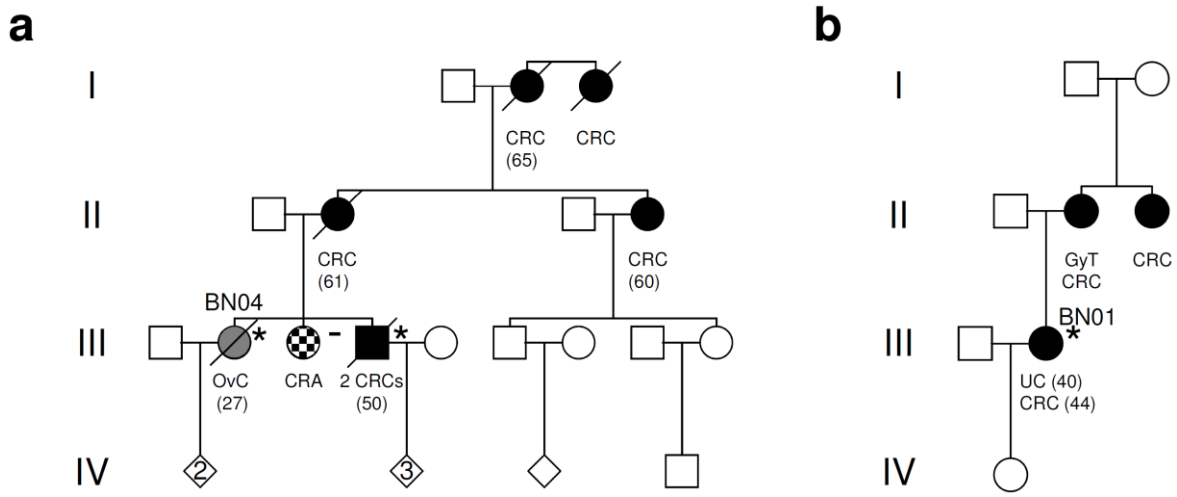


Figure 20. Families with *SEMA4A* G484A (a) and S326F (b) mutations are shown. Asterisk, *SEMA4A* mutation carrier; minus, *SEMA4A* wild-type; black symbol, colorectal cancer; checkered symbol, colorectal adenoma; dark gray, malignant neoplasm; number in symbol, number of unspecified offspring. GyT, gynecologic tumor; OvC, ovarian cancer; UC; uterine cancer. Results of mutational analyses are indicated in tested individuals only. Age at diagnosis is given where known in years in parentheses.

3.5 Database search

At last, it was investigated whether somatically acquired *SEMA4A* mutations are prevalent in colorectal as well as other neoplasms. Analysis of confirmed somatic mutations across different cancer types revealed that *SEMA4A* mutations most frequently occur in cancers from the HNPCC spectrum, i.e. 2.7% (15/559) of colorectal, 2.8% (6/212) of stomach and 3.3% (8/241) of uterine cancers (Alexandrov et al. 2013). In 92% of them, they constitute missense mutations (Figure 6**,** Table 12). In the vast majority of sporadic human neoplasms, *SEMA4A* is affected by gene amplification (Figure 21).

Table 12. Validated somatic *SEMA4A* amino acid altering variants in human cancers. Variants from human cancers could be identified in 2.2% (3/136) of bladder, 0.6% (6/963) of breast, 5.3% (2/38) of cervical, 2.7% (15/559) of colorectal, 1.4% (2/146) of esophageal, 0.8% (3/380) of head, 1.1% (2/176) of lung squamous, 2.5% (10/396) of melanoma, 0.2% (1/471) of ovary, 0.9% (3/330) of prostate, 2.8% (6/212) of stomach and 3.3% (8/241) of uterine cancers, respectively. Ninety-two percent of variants are missense. Additional information was extracted for CRCs. Data refers to (Alexandrov et al. 2013). MSI-L, low MSI; MSI-H, high MSI; *SEMA4A* mRNA, NM_001193300.1; *SEMA4A* protein, NP_001180229.1, n\a, not available.

Tissue	*Sample ID*	*Transcript*	*Protein*	*MutationTaster*	*PP2 HumDiv*	*PP2 HumVar*	*SIFT (cutoff= 0.05)*	*MSI status*	*APC*	*BRAF*	*PIK3CA*	*KRAS*	*TP53*	*Anatomic site*	*Histology*
Bladder	*TCGA-G2-A2EO-01A-11D-A17V-08*	*c.193G>C*	*p.Asp65His*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Bladder	*TCGA-DK-A1A3-01A-11D-A13W-08*	*c.330G>C*	*p.Lys110Asn*	*Polymorphism*	*Probably damaging*	*Probably damaging*	*Damaging*								
Bladder	*TCGA-GC-A3OO-01A-11D-A22Z-08*	*c.1580C>G*	*p.Ser527Cys*	*Polymorphism*	*Probably damaging*	*Possibly damaging*	*Tolerated*								
Breast	*TCGA-AN-A0AT-01A-11D-A045-09*	*c.112C>T*	*p.Pro38Ser*	*Polymorphism*	*Probably damaging*	*Probably damaging*	*Tolerated*								
Breast	*SA075*	*c.486G>T*	*p.Leu162Phe*	*Polymorphism*	*Possibly damaging*	*Benign*	*Tolerated*								
Breast	*PD4980a*	*c.501C>A*	*p.Asp167Glu*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Breast	*TCGA-C8-A274-01A-11D-A16D-09*	*c.1294C>G*	*p.Leu432Val*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Breast	*TCGA-C8-A274-01A-11D-A16D-09*	*c.1867C>T*	*p.Gln623**	*Disease causing*	*n\a*	*n\a*	*n\a*								
Breast	*TCGA-AN-A0FK-01A-11W-A050-09*	*c.2067_2072del*	*p.Leu690_ Phe691del*	*Disease causing*	*n\a*	*n\a*	*n\a*								
Cervix	*TCGA-DR-A0ZM-01A-12D-A10S-08*	*c.960C>G*	*p.Ile320Met*	*Disease causing*	*Possibly damaging*	*Possibly damaging*	*Damaging*								
Cervix	*TCGA-FU-A23L-01A-11D-A16O-08*	*c.1078G>A*	*p.Glu360Lys*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								

Colorectum	TCGA-AZ-6608-01A-11D-1835-10	c.160A>T	p.Ser54Cys	Polymorphism	Possibly damaging	Possibly damaging	Damaging	MSI-L	R186*, Arg1331*	No	No	E62G	P153Afs*28	n\a	Colon adenocarcinoma
Colorectum	TCGA-AA-A00K-01A-02W-A005-10	c.224_225delCT	p.Tyr77Argfs*30	Disease causing	n\a	n\a	n\a	MSI-L	Q1349*	No	No	No	No	n\a	Colon adenocarcinoma
Colorectum	TCGA-DM-A1HB-01A-21D-A183-10	c.338G>A	p.Cys113Tyr	Disease causing	Probably damaging	Probably damaging	Damaging	MSI-H	R499L, S1272*, P2540S	No	No	No	I255T, V218A	Ascending colon	Colon mucinous adenocarcinoma
Colorectum	TCGA-CK-5913-01A-11D-1650-10	c.710T>C	p.Ile237Thr	Disease causing	Benign	Benign	Damaging	MSI-H	S940L, D1425G	L475R	No	No	No	Sigmoid colon	Colon adenocarcinoma
Colorectum	TCGA-AY-6197-01A-11D-1719-10	c.754A>G	p.Ser252Gly	Polymorphism	Possibly damaging	Benign	Tolerated	MSI-H	S1501Ffs*13, R564*, T1705P	No	K986E	No	R175H	Sigmoid colon	Colon adenocarcinoma
Colorectum	TCGA-CM-5861-01A-01D-1650-10	c.833T>G	p.Leu278Arg	Disease causing	Probably damaging	Probably damaging	Damaging	MSI-H	M44V	No	K179R, E365G	No	No	Ascending colon	Colon adenocarcinoma
Colorectum	TCGA-CK-5915-01A-11D-1650-10	c.916C>T	p.His306Tyr	Disease causing	Probably damaging	Probably damaging	Damaging	MSS	S1202P, C1387R, Q1541*, E1552G	No	No	No	No	Cecum	Colon adenocarcinoma
Colorectum	TCGA-AA-3950-01A-02W-0995-10	c.1142T>C	p.Val381Ala	Polymorphism	Benign	Benign	Tolerated	MSI-H	S1465Wfs*3	G615R	S405Y, E522V, G1050S	No	No	Sigmoid colon	Colon adenocarcinoma
Colorectum	TCGA-D5-6928-01A-11D-1924-10	c.1185T>A	p.His395Gln	Disease causing	Probably damaging	Probably damaging	Damaging	MSI-H	S1315P	No	No	No	No	Cecum	Colon adenocarcinoma
Colorectum	TCGA-AD-6889-01A-11D-1924-10	c.1291C>A	p.His431Asn	Polymorphism	Possibly damaging	Benign	Damaging	MSI-H	No	No	No	T20A	No	Sigmoid colon	Colon adenocarcinoma
Colorectum	587232 (doi:10.1038/nature11282)	c.1525G>T	p.Ala509Ser	Disease causing	Probably damaging	Probably damaging	Tolerated	MSS	No	No	C604R	No	No	Cecum	Colon adenocarcinoma
Colorectum	TCGA-CK-5916-01A-11D-1650-10	c.1685C>T	p.Pro562Leu	Disease causing	Possibly damaging	Benign	Damaging	MSI-H	R380L, S1278P, S1495G	No	N1044S	No	No	Cecum	Colon adenocarcinoma
Colorectum	587386 (doi:10.1038/nature11282)	c.1972C>T	p.Arg658Trp	Disease causing	Probably damaging	Possibly damaging	Damaging	MSS	No	No	No	No	No	Ascending colon	Colon mucinous adenocarcinoma
Colorectum	TCGA-DC-5337-01A-01D-1657-10	c.1976A>G	p.Glu659Gly	Disease causing	Benign	Benign	Tolerated	n\a	No	No	No	No	No	n\a	n\a
Colorectum	TCGA-AA-3984-01A-02W-0995-10	c.2230A>G	p.Thr744Ala	Polymorphism	Benign	Benign	Damaging	MSS	F2784C	No	No	No	No	Transverse colon	Colon mucinous adenocarcinoma
Esophagus	ESO-173-Tumor	c.742G>A	p.Glu248Lys	Polymorphism	Possibly damaging	benign	Damaging								
Esophagus	ESO-1488-Tumor	c.1087C>T	p.Arg363Cys	Disease causing	Probably damaging	Probably damaging	Damaging								

Head&Neck	*TCGA-CN-6013-01A-11D-1683-08*	*c.410C>T*	*p.Thr137Ile*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Head&Neck	*TCGA-CV-7101-01A-11D-2012-08*	*c.1283G>A*	*p.Gly428Glu*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Head&Neck	*TCGA-BB-4223-01A-01D-1434-08*	*c.1310G>A*	*p.Gly437Glu*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Lung Adeno	*LUAD-F00365-Tumor*	*c.239C>T*	*p.Ala80Val*	*Disease causing*	*Possibly damaging*	*Possibly damaging*	*Damaging*								
Lung Squamous	*TCGA-34-5231-01A-21D-1817-08*	*c.1412A>G*	*p.Asn471Ser*	*Disease causing*	*benign*	*Benign*	*Tolerated*								
Lung Squamous	*TCGA-18-3410-01A-01D-0983-08*	*c.1973G>T*	*p.Arg658Leu*	*Disease causing*	*benign*	*Benign*	*Damaging*								
Melanoma	*TCGA-FS-A1ZQ-06A-11D-A197-08*	*c.695C>T*	*p.Ser232Phe*	*Disease causing*	*Probably damaging*	*Possibly damaging*	*Damaging*								
Melanoma	*TCGA-EE-A29L-06A-12D-A196-08*	*c.988G>A*	*p.Val330Ile*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Melanoma	*SKCM-Ma-Mel-114-Tumor*	*c.1085C>T*	*p.Ser362Leu*	*Disease causing*	*benign*	*Benign*	*Tolerated*								
Melanoma	*TCGA-EE-A2MF-06A-11D-A21A-08*	*c.1753G>A*	*p.Ala585Thr*	*Disease causing*	*benign*	*Benign*	*Tolerated*								
Melanoma	*TCGA-IH-A3EA-01A-11D-A20D-08*	*c.1826C>T*	*p.Ser609Phe*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Melanoma	*TCGA-EE-A2MR-06A-11D-A196-08*	*c.1849G>A*	*p.Gly617Arg*	*Polymorphism*	*Possibly damaging*	*Benign*	*Tolerated*								
Melanoma	*SKCM-Ma-Mel-94-Tumor*	*c.1889G>A*	*p.Gly630Asp*	*Disease causing*	*benign*	*Benign*	*Tolerated*								
Melanoma	*ME037-Tumor*	*c.1910C>T*	*p.Ser637Phe*	*Polymorphism*	*Probably damaging*	*Probably damaging*	*Tolerated*								
Melanoma	*PD9028a*	*c.2104C>T*	*p.Leu702Phe*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Melanoma	*TCGA-EE-A29C-06A-21D-A197-08*	*c.2117C>T*	*p.Pro706Leu*	*Disease causing*	*benign*	*Benign*	*Tolerated*								
Ovary	*TCGA-36-1568-01A-01W-0615-10*	*c.955C>T*	*p.His319Tyr*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								

Prostate	*PR-1024-Tumor*	*c.785C>T*	*p.Thr262Ile*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Prostate	*TCGA-CH-5771-01A-21D-1576-08*	*c.826G>A*	*p.Glu276Lys*	*Disease causing*	*Possibly damaging*	*Possibly damaging*	*Damaging*								
Prostate	*PR-00-1823-Tumor*	*c.1121C>G*	*p.Pro374Arg*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Stomach	*TCGA-CD-5813-01A-11D-1600-08*	*c.464A>T*	*p.Glu155Val*	*Disease causing*	*Probably damaging*	*Possibly damaging*	*Tolerated*								
Stomach	*TCGA-BR-4298-01A-01D-1126-08*	*c.779T>C*	*p.Leu260Pro*	*Disease causing*	*Probably damaging*	*Possibly damaging*	*Damaging*								
Stomach	*TCGA-BR-7703-01A-11D-2053-08*	*c.1204G>C*	*p.Val402Leu*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Stomach	*TCGA-BR-4200-01A-01D-1126-08*	*c.1514A>G*	*p.Asp505Gly*	*Disease causing*	*Probably damaging*	*Possibly damaging*	*Damaging*								
Stomach	*TCGA-CG-4305-01A-01D-1158-08*	*c.1986G>T*	*p.Lys662Asn*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Stomach	*TCGA-BR-4256-01A-01D-1126-08*	*c.2191A>G*	*p.Ser731Gly*	*Disease causing*	*Possibly damaging*	*Possibly damaging*	*Tolerated*								
Uterus	*TCGA-AP-A059-01A-21D-A122-09*	*c.64C>A*	*p.Leu22Ile*	*Polymorphism*	*benign*	*Benign*	*Tolerated*								
Uterus	*TCGA-AP-A0LT-01A-11W-A062-09*	*c.913C>T*	*p.Arg305Cys*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Uterus	*TCGA-D1-A163-01A-11D-A12J-09*	*c.1087C>T*	*p.Arg363Cys*	*Disease causing*	*Probably damaging*	*Probably damaging*	*Damaging*								
Uterus	*TCGA-D1-A17B-01A-22D-A12J-09*	*c.1134+1G>T*	*p.?*	*Disease causing*	*n\a*	*n\a*	*n\a*								
Uterus	*TCGA-AP-A054-01A-11W-A062-09*	*c.1300A>G*	*p.Met434Val*	*Disease causing*	*benign*	*Benign*	*Damaging*								
Uterus	*TCGA-BS-A0UV-01A-11D-A10B-09*	*c.1408C>T*	*p.Arg470Cys*	*Polymorphism*	*Probably damaging*	*Possibly damaging*	*Damaging*								
Uterus	*TCGA-AX-A0IS-01A-12D-A10M-09*	*c.1481dupG*	*p.Ala495Serfs*4*	*Disease causing*	*n\a*	*n\a*	*n\a*								
Uterus	*TCGA-AP-A0LE-01A-11D-A127-09*	*c.1669C>T*	*p.Arg557Trp*	*Polymorphism*	*Probably damaging*	*Possibly damaging*	*Damaging*								

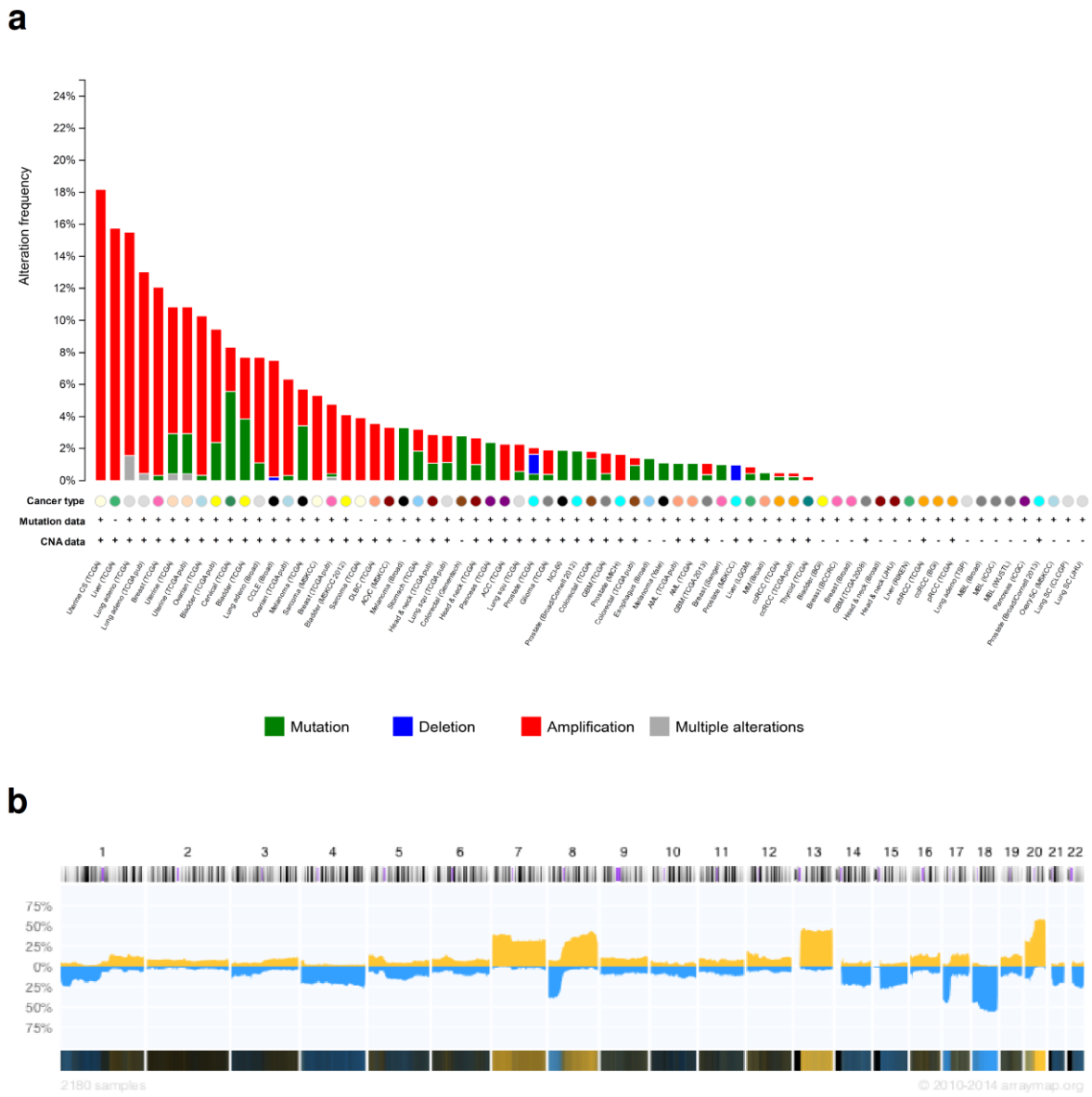


Figure 21. (a) Cross cancer alterations of *SEMA4A*. Data was extracted from cBioPortal for Cancer Genomics (http://www.cbioportal.org/public-portal/) on 24. March 2014. (b) Merged aCGH data of 2180 CRC samples with matched criteria from 34 publications showing gains of chromosome 1q in more than 10% of cases. Data stems from Progenetics (www.progenetics.com) extracted on 26. June 2014.

4 DISCUSSION

The present thesis set out to identify the genetic cause of FCRCX in the Austrian Family K where a monogenetic trait was hypothesized. If identification of a germline variant in a potential novel cancer gene in Family K was successful, this finding should also be corroborated by functional studies and determination of germline variant prevalence in different FCRCX individuals. The aim was to provide evidence for a colorectal cancer predisposing role of identified gene which could eventually result in future surveillance programs. Given the predefined purpose of this study and the presented generated results, it must be concluded that the aims were successfully met.

It has been more than 20 years that novel genes have been identified predisposing to the autosomal dominantly inherited CRC syndrome HNPCC (Stoffel, Eng 2014). During that time, it has been more and more appreciated that a certain overlap of MMR proficient HNPCC with other distinct CRC syndromes does exist as has been shown for *POLE* and *BMPR1A* germline mutations predisposing to polymerase proofreading-associated polyposis and juvenile polyposis syndrome, respectively, so that comprehensive genetic testing becomes more important (Rohlin et al. 2014, Nieminen et al. 2011). The difficulty of identifying the genetic cause in the "basket" of MMR proficient CRCs stems from the heterogeneity of the remaining cases (Jasperson et al. 2010). It is believed that FCRCX is comprised of rare, highly penetrant, monogenic traits on the one hand and of more frequent but less penetrant polygenic features accentuated by environmental factors on the other hand (Jasperson et al. 2010, Stoffel, Eng 2014). For instance, only very recently and simultaneously with publication of the results of our work, a truncating germline mutation in small ribosomal subunit *RPS20* has been strongly associated with a single family out of 26 FCRCX families tested (Nieminen et al. 2014). That and other studies including this thesis about *SEMA4A* highlight the advantages of NGS but also emphasize the importance of well characterized families and phenotypes to find the "needles in the haystack" (Palles et al. 2013, Stoffel, Eng 2014, Nieminen et al. 2014).

This thesis which has begun with one single Austrian FCRCX family shows for the first time that *SEMA4A* germline variants predispose to a hereditary neoplastic syndrome compatible with HNPCC. Two approaches, one a combination of LA and WES, the other solely based on WES, have led to the identification of the *SEMA4A* V78M germline mutation in Family K. This mutation being autosomal dominantly inherited as preconditioned for HNPCC showed incomplete penetrance which was not surprising since cumulative CRC risk is about 15% to 70% in LS depending on the affected MMR gene and even lower in FCRCX (Hampel et al. 2005b, Stoffel et al. 2009, Senter et al. 2008, Ten Broeke et al. 2014, Mueller-Koch et al. 2005, Lindor 2009).

Incomplete penetrance might be a frequent obstacle in the research of FCRCX making definition as affected and unaffected individuals during mutation identification difficult. Explanations in disparity include additional genetic, environmental or life style modifiers playing a role in tumor development (Lynch et al. 2014). Based on hypothetical considerations, the heterozygous *MUTYH* R217H germline variant found in two *SEMA4A* V78M carriers with CRC (K13, K18) could be such a genetic modifier. Biallelic germline *MUTYH* mutations – primarily Y179C and G396D – are the cause of *MUTYH* associated polyposis which is a rare autosomal recessive syndrome with variability in phenotypic presentation resembling FAP (Farrington et al. 2005, Croitoru et al. 2004). Monoallelic *MUTYH* germline variants are associated with a small increased risk of CRC, however, this assumption should be carefully considered since different studies have come to inconsistent results (Farrington et al. 2005, Win et al. 2014, Webb, Rudd & Houlston 2006). The *MUTYH* variant R217H found in K13 and K18 has been previously described once in a cohort of 406 patients with >5 polyps and/or CRC from France but its predisposing role has not been established so far (so called unclassified variant) (Olschwang et al. 2007).

Some individuals with *SEMA4A* V78M germline mutation also developed extracolorectal neoplasms whose risks are considered not to be increased in FCRCX when this entity is seen as a whole (Lindor et al. 2005). Since FCRCX is assumed to be heterogeneous, these extracolorectal tumors, especially breast cancer found in two individuals (K13, K14), might be an integral part of the FCRCX syndrome caused by *SEMA4A* mutations in general or the V78M substitution in particular but more families have to be studied to give a conclusive statement. This is also true for the other *SEMA4A* germline mutations found in this study.

In a mutational screening of 53 FCRCX patients aimed at determining the prevalence of *SEMA4A* germline mutations, two further variants could be identified. In both pedigrees, variants were again associated with extracolorectal neoplasms (ovarian cancer in G484A and endometrial cancer in S326F). It should be noted that ovarian cancer also constitutes part of the HNPCC spectrum (Lynch et al. 2014, Malander et al. 2006). Unfortunately, only the segregation of G484A which followed dominant inheritance could be studied. The homozygous genotype of S326F could be a result of an additional germline mutation unrelated to the other familial cancer cases or may indicate an autosomal recessive mode of inheritance operational in this family. Lack of DNAs from other family members precluded resolution of this question.
The genetic association study of the *SEMA4A* P682S variant found in a higher percentage than expected in tested FCRCX patients showed increased risk of CRC in Austrian and German individuals. This finding should be replicated in an independent cohort with the aim of ascertaining possible ethnic differences. The data suggest that P682S constitutes for a small portion of CRCs probably missed by genome wide association studies detecting mostly frequent, very low penetrant susceptibility loci (Jasperson et al. 2010). It is plausible that also in individuals with P682S further genetic modifiers are operational leading to phenotypic presentation of FCRCX because P682S can also be found in healthy individuals.

SEMA4A has not been implicated in tumor predisposition yet. Compound heterozygous germline *SEMA4A* variants D345H and F350C have been first described in patients with retinitis pigmentosa and cone rode dystrophy but until now this finding has not been replicated (Abid et al. 2006, Ismail et al. 2006). Sema4A-deficient mice exhibit photoreceptor degeneration and disturbed T helper cell function but lack apparently increased tumor development (Rice et al. 2004, Kumanogoh et al. 2005). Given the wide expression of *SEMA4A* in different tissues, it is plausible that mutations might 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 Sema4A-deficient mice in a homozygous knock-in mouse model, a genotype not described in humans (Nojima et al. 2013). This observation stresses the special role of the F350 residue for photoreceptor function and suggests that *SEMA4A* might not be solely responsible for development of retinal diseases in humans

(Nojima et al. 2013). The fact that Sema4A-deficient mice do not develop tumors would not necessarily speak against a possible predisposing role on condition that *SEMA4A* was a true tumor suppressor. First, these mice have not been thoroughly investigated in this direction, and second, deficiencies of human cancer susceptibility gene homologs do not always result in increased tumor formation in mice as has been shown for the MMR gene *Pms2* as well as for *Smad4* predisposing to LS and JPS, respectively (Baker et al. 1998, Takaku et al. 1998). For both genes, additional germline truncating mutations in the gatekeeper gene *Apc* predisposing to FAP in humans are needed for intestinal tumor development in mice.

Germline and somatic *SEMA4A* variants found in this study were not restricted to a certain hot spot supporting a loss-of-function rather than a gain-of-function mechanism (Vogelstein et al. 2013). This assumption is supported by the 7-AAD/BrdU assay performed in the SEMA4A deficient CRC cell line HCT-116 where inhibition was induced by transfection of $SEMA4A^{WT}$ but not $SEMA4A^{V78M}$. Consequently, $SEMA4A^{V78M}$ was unable to down-regulate PI3K/Akt and MAPK/Erk pathways if compared to $SEMA4A^{WT}$.

LOH of V78M was seen in two CRCs from Family K which is a characteristic mechanism of tumor suppressor inactivation (Hanahan, Weinberg 2011). On the other hand, aCGH results of these CRCs show a gain of the region where *SEMA4A* is located, and in the majority of sporadic human neoplasms from public databases, this region is affected by gene amplification hinting at an oncogenic character. There are two possible explanations for this discrepancy. Firstly, *SEMA4A* might exhibit a context specific function in CRCs. It has been recently shown that Sema4A at high levels is able to suppress cell death induced by plexin D1 (Plxnd1) in the mouse mammary tumor cell line 4T1 whereas the same semaphorin-plexin constellation inhibited proliferation in human endothelial cells supporting the concept of context specificity with two completely opposing cellular processes (Luchino et al. 2013, Toyofuku et al. 2007). Secondly, it might be that loss of the *SEMA4A* wild-type allele accompanied by amplification of the mutant one was the – although unusual – mechanism of tumor suppressor inactivation in this particular entity. Specimens from patients with FCRCX frequently exhibit gains of different chromosomal regions including chromosome 1 which is accompanied by copy-neutral LOH (Middeldorp et

al. 2012). One hypothetical explanation of amplification might be the presence of multiple oncogenes on chromosome 1q which is suggested by frequent gains found in more than 10% of over 2000 CRCs analyzed. Unfortunately, due to lack of good quality DNA from FFPE tumor specimens, we were unable to perform SNP arrays to prove the possibility of copy-neutral LOH. However, we were able to perform quantitative allele specific dPCR in CRCs of K13 and K26. This analysis confirmed the results from direct sequencing.

Taken together, at the moment, *SEMA4A* can neither be unambiguously defined as a tumor suppressor acting through second hits and/or haploinsufficiency nor as an oncogene acting through neomorphic gain-of-function so that this question remains to be answered. Such difficulties in classification are not new and a hybrid-mechanism similar to *TP53* where loss of function, gain of function, dominant negative effect and haploinsufficiency have been described for the same protein is conceivable (Muller, Vousden 2014).

The *in vitro* results have shown that SEMA4A^{V78M} differentially modulates the PI3K/Akt and MAPK/Erk pathways in HCT-116 cells, and that additional molecular hits are probably needed to form a malignant phenotype since 293T showed no effects on these pathways after transfection. Additional hits could be for instance mutations in the *PIK3CA* and *KRAS* genes as found in the HCT-116 cell line (H1047R [*PIC3CA*], G13D [*KRAS*], according to data from Broad-Novartis Cancer Cell Line Encyclopedia, http://www.broadinstitute.org/ccle/home, extracted March 2014) affecting these oncogenic pathways or differential expression of plexin and other yet unidentified receptors (Samuels et al. 2005, Caron et al. 2005). Indeed, co-operating mutations in known driver genes could be detected by targeted deep sequencing of CRCs of V78M carriers. However, this is a mere description which can be helpful for future investigations on this topic. The author is aware that the low number of tumors investigated precludes any conclusion on a firm association between the germline variant and the somatic mutations observed. Nevertheless, this finding is in accordance with well-established concepts of predisposing germline mutations where further somatically acquired events in other tumor-associated genes are necessary for manifestation of the malignant phenotype (Vogelstein et al. 2013).

Therefore, it is hypothesized, based on available data, that CRCs in V78M carriers probably followed the conventional route of colorectal carcinomagenesis.

Given the pleiotropic expression of *SEMA4A* in different tissues including immune and endothelial cells, one can speculate about an additional role in the vascular microenvironment of colonic tumors and the immune response against them (Kumanogoh et al. 2005, Toyofuku et al. 2007, Kumanogoh et al. 2002, Delgoffe et al. 2013). Although this thesis did not investigate in this direction, the author acknowledges that further research covering this area is warranted.
Another important question that had not been addressed in this thesis is the mechanism by which *SEMA4A* mutations act on the molecular level. SEMA4A is a membrane-bound ligand that can bind plexins B1, B2, B3 and D1 as well as TIM2 (Toyofuku et al. 2007, Kumanogoh et al. 2002). It can be proteolyticly cleaved because it has been detected at increased levels in the sera of patients with multiple sclerosis (Nakatsuji et al. 2012). Therefore, one must assume that SEMA4A can not only work by direct cell to cell binding but also in autocrine and paracrine manners. Mutations in *SEMA4A* might disturb these cell-cell-interactions either by increasing or decreasing activities of subsequent cell signaling pathways. Whether reverse signaling through the cytoplasmic portion of SEMA4A exits, is currently unknown but seems unlikely given the short cytoplasmic tail without known protein domains. As a consequence, the most important next objective should be the identification of relevant binding partners of SEMA4A in CRC if one wants to understand how *SEMA4A* mutations mechanistically work.

In summary, the data presented in this thesis provide good evidence for a CRC predisposition role of *SEMA4A* broadening the little yet growing knowledge about the pathophysiological role of semaphorins in human carcinogenesis. Still, regarding *SEMA4A*, there is much work ahead for both, biologists and clinicians: at the bench by investigating the exact mechanism thereby paving the way for potential new therapeutics, but also in the clinics by defining appropriate screening measures for early tumor detection of mutation carriers leading to reduced morbidity and mortality in patients as well as family members.

5 References

Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. 2002, "Merlin--rapid analysis of dense genetic maps using sparse gene flow trees", *Nature genetics,* vol. 30, no. 1, pp. 97-101.

Abid, A., Ismail, M., Mehdi, S.Q. & Khaliq, S. 2006, "Identification of novel mutations in the SEMA4A gene associated with retinal degenerative diseases", *Journal of medical genetics,* vol. 43, no. 4, pp. 378-381.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjord, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jager, N., Jones, D.T., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., Lopez-Otin, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N., Valdes-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Zucman-Rossi, J., Futreal, P.A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J. & Stratton, M.R. 2013, "Signatures of mutational processes in human cancer", *Nature,* vol. 500, no. 7463, pp. 415-421.

Bahadori, B., Uitz, E., Mayer, A., Harauer, J., Dam, K., Truschnig-Wilders, M., Pilger, E. & Renner, W. 2010, "Polymorphisms of the hypoxia-inducible factor 1 gene and peripheral artery disease", *Vascular medicine (London, England),* vol. 15, no. 5, pp. 371-374.

Baker, S.M., Harris, A.C., Tsao, J.L., Flath, T.J., Bronner, C.E., Gordon, M., Shibata, D. & Liskay, R.M. 1998, "Enhanced intestinal adenomatous polyp formation in Pms2-/-;Min mice", *Cancer research,* vol. 58, no. 6, pp. 1087-1089.

Balmana, J., Balaguer, F., Cervantes, A., Arnold, D. & ESMO Guidelines Working Group 2013, "Familial risk-colorectal cancer: ESMO Clinical Practice Guidelines", *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO,* vol. 24 Suppl 6, pp. vi73-80.

Beillard, E., Pallisgaard, N., van der Velden, V.H., Bi, W., Dee, R., van der Schoot, E., Delabesse, E., Macintyre, E., Gottardi, E., Saglio, G., Watzinger, F., Lion, T., van Dongen, J.J., Hokland, P. & Gabert, J. 2003, "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 : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K,* vol. 17, no. 12, pp. 2474-2486.

Boland, C.R. & Troncale, F.J. 1984, "Familial colonic cancer without antecedent polyposis", *Annals of Internal Medicine,* vol. 100, no. 5, pp. 700-701.

Bonadona, V., Bonaiti, B., Olschwang, S., Grandjouan, S., Huiart, L., Longy, M., Guimbaud, R., Buecher, B., Bignon, Y.J., Caron, O., Colas, C., Nogues, C., Lejeune-Dumoulin, S., Olivier-Faivre, L., Polycarpe-Osaer, F., Nguyen, T.D., Desseigne, F., Saurin, J.C., Berthet, P., Leroux, D., Duffour, J., Manouvrier, S., Frebourg, T., Sobol, H., Lasset, C., Bonaiti-Pellie, C. & French Cancer Genetics Network 2011, "Cancer

risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome", *Jama,* vol. 305, no. 22, pp. 2304-2310.
Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J. & Lindblom, A. 1994, "Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer", *Nature,* vol. 368, no. 6468, pp. 258-261.
Brosens, L.A., van Hattem, A., Hylind, L.M., Iacobuzio-Donahue, C., Romans, K.E., Axilbund, J., Cruz-Correa, M., Tersmette, A.C., Offerhaus, G.J. & Giardiello, F.M. 2007, "Risk of colorectal cancer in juvenile polyposis", *Gut,* vol. 56, no. 7, pp. 965-967.
Cagnoni, G. & Tamagnone, L. 2013, "Semaphorin receptors meet receptor tyrosine kinases on the way of tumor progression", *Oncogene,* vol. 33, no. 40, pp. 4795-4802.
Cancer Genome Atlas Network 2012, "Comprehensive molecular characterization of human colon and rectal cancer", *Nature,* vol. 487, no. 7407, pp. 330-337.
Caron, R.W., Yacoub, A., Li, M., Zhu, X., Mitchell, C., Hong, Y., Hawkins, W., Sasazuki, T., Shirasawa, S., Kozikowski, A.P., Dennis, P.A., Hagan, M.P., Grant, S. & Dent, P. 2005, "Activated forms of H-RAS and K-RAS differentially regulate membrane association of PI3K, PDK-1, and AKT and the effect of therapeutic kinase inhibitors on cell survival", *Molecular cancer therapeutics,* vol. 4, no. 2, pp. 257-270.
Chan, T.L., Yuen, S.T., Kong, C.K., Chan, Y.W., Chan, A.S., Ng, W.F., Tsui, W.Y., Lo, M.W., Tam, W.Y., Li, V.S. & Leung, S.Y. 2006, "Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer", *Nature genetics,* vol. 38, no. 10, pp. 1178-1183.
Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. 2012, "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,* vol. 6, no. 2, pp. 80-92.
Croitoru, M.E., Cleary, S.P., Di Nicola, N., Manno, M., Selander, T., Aronson, M., Redston, M., Cotterchio, M., Knight, J., Gryfe, R. & Gallinger, S. 2004, "Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk", *Journal of the National Cancer Institute,* vol. 96, no. 21, pp. 1631-1634.
Delgoffe, G.M., Woo, S.R., Turnis, M.E., Gravano, D.M., Guy, C., Overacre, A.E., Bettini, M.L., Vogel, P., Finkelstein, D., Bonnevier, J., Workman, C.J. & Vignali, D.A. 2013, "Stability and function of regulatory T cells is maintained by a neuropilin-1-semaphorin-4a axis", *Nature,* vol. 501, no. 7466, pp. 252-256.
Deng, G., Bell, I., Crawley, S., Gum, J., Terdiman, J.P., Allen, B.A., Truta, B., Sleisenger, M.H. & Kim, Y.S. 2004, "BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 10, no. 1 Pt 1, pp. 191-195.
DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. & Daly, M.J. 2011, "A framework for variation discovery and genotyping using next-generation DNA sequencing data", *Nature genetics,* vol. 43, no. 5, pp. 491-498.
Engel, C., Loeffler, M., Steinke, V., Rahner, N., Holinski-Feder, E., Dietmaier, W., Schackert, H.K., Goergens, H., von Knebel Doeberitz, M., Goecke, T.O., Schmiegel, W., Buettner, R., Moeslein, G., Letteboer, T.G., Gomez Garcia, E., Hes, F.J., Hoogerbrugge, N., Menko, F.H., van Os, T.A., Sijmons, R.H., Wagner, A., Kluijt, I.,

Propping, P. & Vasen, H.F. 2012, "Risks of less common cancers in proven mutation carriers with lynch syndrome", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 30, no. 35, pp. 4409-4415.
Farrington, S.M., Tenesa, A., Barnetson, R., Wiltshire, A., Prendergast, J., Porteous, M., Campbell, H. & Dunlop, M.G. 2005, "Germline susceptibility to colorectal cancer due to base-excision repair gene defects", *American Journal of Human Genetics,* vol. 77, no. 1, pp. 112-119.
Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W., Comber, H., Forman, D. & Bray, F. 2013, "Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012", *European journal of cancer (Oxford, England : 1990),* vol. 49, no. 6, pp. 1374-1403.
Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M. & Kolodner, R. 1993, "The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer", *Cell,* vol. 75, no. 5, pp. 1027-1038.
Foulkes, W.D. 2008, "Inherited susceptibility to common cancers", *The New England journal of medicine,* vol. 359, no. 20, pp. 2143-2153.
Hampel, H., Frankel, W.L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Clendenning, M., Sotamaa, K., Prior, T., Westman, J.A., Panescu, J., Fix, D., Lockman, J., LaJeunesse, J., Comeras, I. & de la Chapelle, A. 2008, "Feasibility of screening for Lynch syndrome among patients with colorectal cancer", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 26, no. 35, pp. 5783-5788.
Hampel, H., Frankel, W.L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Nakagawa, H., Sotamaa, K., Prior, T.W., Westman, J., Panescu, J., Fix, D., Lockman, J., Comeras, I. & de la Chapelle, A. 2005a, "Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer)", *The New England journal of medicine,* vol. 352, no. 18, pp. 1851-1860.
Hampel, H., Stephens, J.A., Pukkala, E., Sankila, R., Aaltonen, L.A., Mecklin, J.P. & de la Chapelle, A. 2005b, "Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset", *Gastroenterology,* vol. 129, no. 2, pp. 415-421.
Hanahan, D. & Weinberg, R.A. 2011, "Hallmarks of cancer: the next generation", *Cell,* vol. 144, no. 5, pp. 646-674.
Hearle, N., Schumacher, V., Menko, F.H., Olschwang, S., Boardman, L.A., Gille, J.J., Keller, J.J., Westerman, A.M., Scott, R.J., Lim, W., Trimbath, J.D., Giardiello, F.M., Gruber, S.B., Offerhaus, G.J., de Rooij, F.W., Wilson, J.H., Hansmann, A., Moslein, G., Royer-Pokora, B., Vogel, T., Phillips, R.K., Spigelman, A.D. & Houlston, R.S. 2006, "Frequency and spectrum of cancers in the Peutz-Jeghers syndrome", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 12, no. 10, pp. 3209-3215.
Hitchins, M.P., Rapkins, R.W., Kwok, C.T., Srivastava, S., Wong, J.J., Khachigian, L.M., Polly, P., Goldblatt, J. & Ward, R.L. 2011, "Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR", *Cancer cell,* vol. 20, no. 2, pp. 200-213.
Ismail, M., Abid, A., Anwar, K., Mehdi, S.Q. & Khaliq, S. 2006, "Refinement of the locus for autosomal recessive cone-rod dystrophy (CORD8) linked to chromosome 1q23-q24 in a Pakistani family and exclusion of candidate genes", *Journal of human genetics,* vol. 51, no. 9, pp. 827-831.
Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P.R., Davis, H., Kaur, K., Heinimann, K., Howarth, K., East, J., Taylor, J., Thomas, H. & Tomlinson, I.

2012, "Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1", *Nature genetics,* vol. 44, no. 6, pp. 699-703.
Jahn, S., Minai-Pour, M.B., Speicher, M.R., Reiner-Concin, A. & Hoefler, G. 2009, "Comprehensive screening for Lynch syndrome: who can be the driving force in daily clinical practice?", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 27, no. 13, pp. 2292.
Janssen, B.J., Robinson, R.A., Perez-Branguli, F., Bell, C.H., Mitchell, K.J., Siebold, C. & Jones, E.Y. 2010, "Structural basis of semaphorin-plexin signalling", *Nature,* vol. 467, no. 7319, pp. 1118-1122.
Jasperson, K.W., Tuohy, T.M., Neklason, D.W. & Burt, R.W. 2010, "Hereditary and familial colon cancer", *Gastroenterology,* vol. 138, no. 6, pp. 2044-2058.
Kelley, L.A. & Sternberg, M.J. 2009, "Protein structure prediction on the Web: a case study using the Phyre server", *Nature protocols,* vol. 4, no. 3, pp. 363-371.
Kievit, W., de Bruin, J.H., Adang, E.M., Ligtenberg, M.J., Nagengast, F.M., van Krieken, J.H. & Hoogerbrugge, N. 2004, "Current clinical selection strategies for identification of hereditary non-polyposis colorectal cancer families are inadequate: a meta-analysis", *Clinical genetics,* vol. 65, no. 4, pp. 308-316.
Klarskov, L., Holck, S., Bernstein, I. & Nilbert, M. 2012, "Hereditary colorectal cancer diagnostics: morphological features of familial colorectal cancer type X versus Lynch syndrome", *Journal of clinical pathology,* vol. 65, no. 4, pp. 352-356.
Kumanogoh, A., Marukawa, S., Suzuki, K., Takegahara, N., Watanabe, C., Ch'ng, E., Ishida, I., Fujimura, H., Sakoda, S., Yoshida, K. & Kikutani, H. 2002, "Class IV semaphorin Sema4A enhances T-cell activation and interacts with Tim-2", *Nature,* vol. 419, no. 6907, pp. 629-633.
Kumanogoh, A., Shikina, T., Suzuki, K., Uematsu, S., Yukawa, K., Kashiwamura, S., Tsutsui, H., Yamamoto, M., Takamatsu, H., Ko-Mitamura, E.P., Takegahara, N., Marukawa, S., Ishida, I., Morishita, H., Prasad, D.V., Tamura, M., Mizui, M., Toyofuku, T., Akira, S., Takeda, K., Okabe, M. & Kikutani, H. 2005, "Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice", *Immunity,* vol. 22, no. 3, pp. 305-316.
Lansdorp-Vogelaar, I., Kuntz, K.M., Knudsen, A.B., van Ballegooijen, M., Zauber, A.G. & Jemal, A. 2012, "Contribution of screening and survival differences to racial disparities in colorectal cancer rates", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology,* vol. 21, no. 5, pp. 728-736.
Li, G.M. 2008, "Mechanisms and functions of DNA mismatch repair", *Cell research,* vol. 18, no. 1, pp. 85-98.
Li, H. & Durbin, R. 2010, "Fast and accurate long-read alignment with Burrows-Wheeler transform", *Bioinformatics (Oxford, England),* vol. 26, no. 5, pp. 589-595.
Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. & Hemminki, K. 2000, "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland", *The New England journal of medicine,* vol. 343, no. 2, pp. 78-85.
Ligtenberg, M.J., Kuiper, R.P., Chan, T.L., Goossens, M., Hebeda, K.M., Voorendt, M., Lee, T.Y., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S.J., Tsui, W.Y., Kong, C.K., Brunner, H.G., van Kessel, A.G., Yuen, S.T., van Krieken, J.H., Leung, S.Y. & Hoogerbrugge, N. 2009, "Heritable somatic methylation and inactivation of MSH2

in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1", *Nature genetics,* vol. 41, no. 1, pp. 112-117.
Lindor, N.M. 2009, "Familial colorectal cancer type X: the other half of hereditary nonpolyposis colon cancer syndrome", *Surgical oncology clinics of North America,* vol. 18, no. 4, pp. 637-645.
Lindor, N.M., Rabe, K., Petersen, G.M., Haile, R., Casey, G., Baron, J., Gallinger, S., Bapat, B., Aronson, M., Hopper, J., Jass, J., LeMarchand, L., Grove, J., Potter, J., Newcomb, P., Terdiman, J.P., Conrad, P., Moslein, G., Goldberg, R., Ziogas, A., Anton-Culver, H., de Andrade, M., Siegmund, K., Thibodeau, S.N., Boardman, L.A. & Seminara, D. 2005, "Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X", *JAMA : the journal of the American Medical Association,* vol. 293, no. 16, pp. 1979-1985.
Llor, X., Pons, E., Xicola, R.M., Castells, A., Alenda, C., Pinol, V., Andreu, M., Castellvi-Bel, S., Paya, A., Jover, R., Bessa, X., Giros, A., Roca, A., Gassull, M.A. & Gastrointestinal Oncology Group of the Spanish Gastroenterological Association 2005, "Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway", *Clinical cancer research : an official journal of the American Association for Cancer Research,* vol. 11, no. 20, pp. 7304-7310.
Lubbe, S.J., Di Bernardo, M.C., Chandler, I.P. & Houlston, R.S. 2009, "Clinical implications of the colorectal cancer risk associated with MUTYH mutation", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 27, no. 24, pp. 3975-3980.
Luchino, J., Hocine, M., Amoureux, M.C., Gibert, B., Bernet, A., Royet, A., Treilleux, I., Lecine, P., Borg, J.P., Mehlen, P., Chauvet, S. & Mann, F. 2013, "Semaphorin 3E suppresses tumor cell death triggered by the plexin D1 dependence receptor in metastatic breast cancers", *Cancer cell,* vol. 24, no. 5, pp. 673-685.
Lynch, H.T. & de la Chapelle, A. 2003, "Hereditary colorectal cancer", *The New England journal of medicine,* vol. 348, no. 10, pp. 919-932.
Lynch, H.T., Drescher, K., Knezetic, J. & Lanspa, S. 2014, "Genetics, biomarkers, hereditary cancer syndrome diagnosis, heterogeneity and treatment: a review", *Current treatment options in oncology,* vol. 15, no. 3, pp. 429-442.
Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F. & Boland, C.R. 2009, "Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications", *Clinical genetics,* vol. 76, no. 1, pp. 1-18.
Lynch, H.T., Shaw, M.W., Magnuson, C.W., Larsen, A.L. & Krush, A.J. 1966, "Hereditary Factors in Cancer - Study of 2 Large Midwestern Kindreds", *Archives of Internal Medicine,* vol. 117, no. 2, pp. 206-212.
Malander, S., Rambech, E., Kristoffersson, U., Halvarsson, B., Ridderheim, M., Borg, A. & Nilbert, M. 2006, "The contribution of the hereditary nonpolyposis colorectal cancer syndrome to the development of ovarian cancer", *Gynecologic oncology,* vol. 101, no. 2, pp. 238-243.
Middeldorp, A., van Eijk, R., Oosting, J., Forte, G.I., van Puijenbroek, M., van Nieuwenhuizen, M., Corver, W.E., Ruano, D., Caldes, T., Wijnen, J., Morreau, H. & van Wezel, T. 2012, "Increased frequency of 20q gain and copy-neutral loss of heterozygosity in mismatch repair proficient familial colorectal carcinomas", *International journal of cancer.Journal international du cancer,* vol. 130, no. 4, pp. 837-846.

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M. & Mori, T. 1997, "Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer", *Nature genetics,* vol. 17, no. 3, pp. 271-272.

Morak, M., Heidenreich, B., Keller, G., Hampel, H., Laner, A., de la Chapelle, A. & Holinski-Feder, E. 2014, "Biallelic MUTYH mutations can mimic Lynch syndrome", *European journal of human genetics : EJHG,* vol. 22, no. 11, pp. 1334-1337.

Moreira, L., Balaguer, F., Lindor, N., de la Chapelle, A., Hampel, H., Aaltonen, L.A., Hopper, J.L., Le Marchand, L., Gallinger, S., Newcomb, P.A., Haile, R., Thibodeau, S.N., Gunawardena, S., Jenkins, M.A., Buchanan, D.D., Potter, J.D., Baron, J.A., Ahnen, D.J., Moreno, V., Andreu, M., Ponz de Leon, M., Rustgi, A.K., Castells, A. & EPICOLON Consortium 2012, "Identification of Lynch syndrome among patients with colorectal cancer", *Jama,* vol. 308, no. 15, pp. 1555-1565.

Mueller-Koch, Y., Vogelsang, H., Kopp, R., Lohse, P., Keller, G., Aust, D., Muders, M., Gross, M., Daum, J., Schiemann, U., Grabowski, M., Scholz, M., Kerker, B., Becker, I., Henke, G. & Holinski-Feder, E. 2005, "Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer", *Gut,* vol. 54, no. 12, pp. 1733-1740.

Muller, P.A. & Vousden, K.H. 2014, "Mutant p53 in Cancer: New Functions and Therapeutic Opportunities", *Cancer cell,* vol. 25, no. 3, pp. 304-317.

Nakatsuji, Y., Okuno, T., Moriya, M., Sugimoto, T., Kinoshita, M., Takamatsu, H., Nojima, S., Kimura, T., Kang, S., Ito, D., Nakagawa, Y., Toyofuku, T., Takata, K., Nakano, M., Kubo, M., Suzuki, S., Matsui-Hasumi, A., Uto-Konomi, A., Ogata, A., Mochizuki, H., Sakoda, S. & Kumanogoh, A. 2012, "Elevation of Sema4A implicates Th cell skewing and the efficacy of IFN-beta therapy in multiple sclerosis", *Journal of immunology (Baltimore, Md.: 1950),* vol. 188, no. 10, pp. 4858-4865.

Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D. & Fraser, C.M. 1994, "Mutations of two PMS homologues in hereditary nonpolyposis colon cancer", *Nature,* vol. 371, no. 6492, pp. 75-80.

Nieminen, T.T., Abdel-Rahman, W.M., Ristimaki, A., Lappalainen, M., Lahermo, P., Mecklin, J.P., Jarvinen, H.J. & Peltomaki, P. 2011, "BMPR1A mutations in hereditary nonpolyposis colorectal cancer without mismatch repair deficiency", *Gastroenterology,* vol. 141, no. 1, pp. e23-26.

Nieminen, T.T., O'Donohue, M.F., Wu, Y., Lohi, H., Scherer, S.W., Paterson, A.D., Ellonen, P., Abdel-Rahman, W.M., Valo, S., Mecklin, J.P., Jarvinen, H.J., Gleizes, P.E. & Peltomaki, P. 2014, "Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency", *Gastroenterology,* vol. 147, no. 3, pp. 595-598.e5.

Nojima, S., Toyofuku, T., Kamao, H., Ishigami, C., Kaneko, J., Okuno, T., Takamatsu, H., Ito, D., Kang, S., Kimura, T., Yoshida, Y., Morimoto, K., Maeda, Y., Ogata, A., Ikawa, M., Morii, E., Aozasa, K., Takagi, J., Takahashi, M. & Kumanogoh, A. 2013, "A point mutation in Semaphorin 4A associates with defective endosomal sorting and causes retinal degeneration", *Nature communications,* vol. 4, pp. 1406.

Olschwang, S., Blanche, H., de Moncuit, C. & Thomas, G. 2007, "Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the MYH gene identified in a population with adenomatous polyposis", *Genetic testing,* vol. 11, no. 3, pp. 315-320.

Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L.G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M.B., Humphray, S., CORGI Consortium, WGS500 Consortium, Lucassen, A., Holmes, C.C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E.J., Kerr, D.J., Clark, S., Grimes, J., Kearsey, S.E., Thomas, H.J., McVean, G., Houlston, R.S. & Tomlinson, I. 2013, "Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas", *Nature genetics,* vol. 45, no. 2, pp. 136-144.

Rice, D.S., Huang, W., Jones, H.A., Hansen, G., Ye, G.L., Xu, N., Wilson, E.A., Troughton, K., Vaddi, K., Newton, R.C., Zambrowicz, B.P. & Sands, A.T. 2004, "Severe retinal degeneration associated with disruption of semaphorin 4A", *Investigative ophthalmology & visual science,* vol. 45, no. 8, pp. 2767-2777.

Rodriguez, S., Gaunt, T.R. & Day, I.N. 2009, "Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies", *American Journal of Epidemiology,* vol. 169, no. 4, pp. 505-514.

Rodriguez-Moranta, F., Castells, A., Andreu, M., Pinol, V., Castellvi-Bel, S., Alenda, C., Llor, X., Xicola, R.M., Jover, R., Paya, A., Bessa, X., Balaguer, F., Cubiella, J., Arguello, L., Morillas, J.D., Bujanda, L. & Gastrointestinal Oncology Group of the Spanish Gastroenterological Association 2006, "Clinical performance of original and revised Bethesda guidelines for the identification of MSH2/MLH1 gene carriers in patients with newly diagnosed colorectal cancer: proposal of a new and simpler set of recommendations", *The American Journal of Gastroenterology,* vol. 101, no. 5, pp. 1104-1111.

Rohlin, A., Zagoras, T., Nilsson, S., Lundstam, U., Wahlstrom, J., Hulten, L., Martinsson, T., Karlsson, G.B. & Nordling, M. 2014, "A mutation in POLE predisposing to a multi-tumour phenotype", *International journal of oncology,* vol. 45, no. 1, pp. 77-81.

Ruschendorf, F. & Nurnberg, P. 2005, "ALOHOMORA: a tool for linkage analysis using 10K SNP array data", *Bioinformatics (Oxford, England),* vol. 21, no. 9, pp. 2123-2125.

Samuels, Y., Diaz, L.A.,Jr, Schmidt-Kittler, O., Cummins, J.M., Delong, L., Cheong, I., Rago, C., Huso, D.L., Lengauer, C., Kinzler, K.W., Vogelstein, B. & Velculescu, V.E. 2005, "Mutant PIK3CA promotes cell growth and invasion of human cancer cells", *Cancer cell,* vol. 7, no. 6, pp. 561-573.

Schulz, E., Valentin, A., Ulz, P., Beham-Schmid, C., Lind, K., Rupp, V., Lackner, H., Wolfler, A., Zebisch, A., Olipitz, W., Geigl, J., Berghold, A., Speicher, M.R. & Sill, H. 2012, "Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms", *Journal of medical genetics,* vol. 149, no. 7, pp. 422-428.

Senter, L., Clendenning, M., Sotamaa, K., Hampel, H., Green, J., Potter, J.D., Lindblom, A., Lagerstedt, K., Thibodeau, S.N., Lindor, N.M., Young, J., Winship, I., Dowty, J.G., White, D.M., Hopper, J.L., Baglietto, L., Jenkins, M.A. & de la Chapelle, A. 2008, "The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations", *Gastroenterology,* vol. 135, no. 2, pp. 419-428.

Siegel, R., Desantis, C. & Jemal, A. 2014, "Colorectal cancer statistics, 2014", *CA: a cancer journal for clinicians,* vol. 64, no. 2, pp. 104-117.

Soravia, C., Berk, T., Madlensky, L., Mitri, A., Cheng, H., Gallinger, S., Cohen, Z. & Bapat, B. 1998, "Genotype-phenotype correlations in attenuated adenomatous polyposis coli", *American Journal of Human Genetics,* vol. 62, no. 6, pp. 1290-1301.

Spier, I., Holzapfel, S., Altmuller, J., Zhao, B., Horpaopan, S., Vogt, S., Chen, S., Morak, M., Raeder, S., Kayser, K., Stienen, D., Adam, R., Nurnberg, P., Plotz, G., Holinski-Feder, E., Lifton, R.P., Thiele, H., Hoffmann, P., Steinke, V. & Aretz, S. 2014, "Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas", *International journal of cancer.Journal international du cancer,* vol. 137, no. 2, pp. 320-331.

Stoffel, E., Mukherjee, B., Raymond, V.M., Tayob, N., Kastrinos, F., Sparr, J., Wang, F., Bandipalliam, P., Syngal, S. & Gruber, S.B. 2009, "Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome", *Gastroenterology,* vol. 137, no. 5, pp. 1621-1627.

Stoffel, E.M. & Eng, C. 2014, "Exome sequencing in familial colorectal cancer: searching for needles in haystacks", *Gastroenterology,* vol. 147, no. 3, pp. 554-556.

Stoffel, E.M., Mangu, P.B., Gruber, S.B., Hamilton, S.R., Kalady, M.F., Lau, M.W., Lu, K.H., Roach, N., Limburg, P.J., American Society of Clinical Oncology & European Society of Clinical Oncology 2015, "Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 33, no. 2, pp. 209-217.

Takaku, K., Oshima, M., Miyoshi, H., Matsui, M., Seldin, M.F. & Taketo, M.M. 1998, "Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes", *Cell,* vol. 92, no. 5, pp. 645-656.

Ten Broeke, S.W., Brohet, R.M., Tops, C.M., van der Klift, H.M., Velthuizen, M.E., Bernstein, I., Capella Munar, G., Gomez Garcia, E., Hoogerbrugge, N., Letteboer, T.G., Menko, F.H., Lindblom, A., Mensenkamp, A.R., Moller, P., van Os, T.A., Rahner, N., Redeker, B.J., Sijmons, R.H., Spruijt, L., Suerink, M., Vos, Y.J., Wagner, A., Hes, F.J., Vasen, H.F., Nielsen, M. & Wijnen, J.T. 2014, "Lynch Syndrome Caused by Germline PMS2 Mutations: Delineating the Cancer Risk", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology,* vol. 33, no. 4, pp. 319-325.

Toyofuku, T., Yabuki, M., Kamei, J., Kamei, M., Makino, N., Kumanogoh, A. & Hori, M. 2007, "Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1", *The EMBO journal,* vol. 26, no. 5, pp. 1373-1384.

Tsuruma, K., Nishimura, Y., Kishi, S., Shimazawa, M., Tanaka, T. & Hara, H. 2012, "SEMA4A mutations lead to susceptibility to light irradiation, oxidative stress, and ER stress in retinal pigment epithelial cells", *Investigative ophthalmology & visual science,* vol. 53, no. 10, pp. 6729-6737.

Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Ruschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., Hamilton, S.R., Hiatt, R.A., Jass, J., Lindblom, A., Lynch, H.T., Peltomaki, P., Ramsey, S.D., Rodriguez-Bigas, M.A., Vasen, H.F., Hawk, E.T., Barrett, J.C., Freedman, A.N. & Srivastava, S. 2004a, "Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability", *Journal of the National Cancer Institute,* vol. 96, no. 4, pp. 261-268.

Umar, A., Risinger, J.I., Hawk, E.T. & Barrett, J.C. 2004b, "Testing guidelines for hereditary non-polyposis colorectal cancer", *Nature reviews.Cancer,* vol. 4, no. 2, pp. 153-158.

Vasen, H.F., Blanco, I., Aktan-Collan, K., Gopie, J.P., Alonso, A., Aretz, S., Bernstein, I., Bertario, L., Burn, J., Capella, G., Colas, C., Engel, C., Frayling, I.M., Genuardi, M., Heinimann, K., Hes, F.J., Hodgson, S.V., Karagiannis, J.A., Lalloo, F., Lindblom, A., Mecklin, J.P., Moller, P., Myrhoj, T., Nagengast, F.M., Parc, Y., Ponz de Leon, M., Renkonen-Sinisalo, L., Sampson, J.R., Stormorken, A., Sijmons, R.H., Tejpar, S., Thomas, H.J., Rahner, N., Wijnen, J.T., Jarvinen, H.J., Moslein, G. & Mallorca group 2013, "Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts", *Gut,* vol. 62, no. 6, pp. 812-823.

Vasen, H.F., Mecklin, J.P., Khan, P.M. & Lynch, H.T. 1991, "The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)", *Diseases of the colon and rectum,* vol. 34, no. 5, pp. 424-425.

Vasen, H.F., Watson, P., Mecklin, J.P. & Lynch, H.T. 1999, "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC", *Gastroenterology,* vol. 116, no. 6, pp. 1453-1456.

Vogelaar, I., van Ballegooijen, M., Schrag, D., Boer, R., Winawer, S.J., Habbema, J.D. & Zauber, A.G. 2006, "How much can current interventions reduce colorectal cancer mortality in the U.S.? Mortality projections for scenarios of risk-factor modification, screening, and treatment", *Cancer,* vol. 107, no. 7, pp. 1624-1633.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A.,Jr & Kinzler, K.W. 2013, "Cancer genome landscapes", *Science (New York, N.Y.),* vol. 339, no. 6127, pp. 1546-1558.

Wang, K., Li, M. & Hakonarson, H. 2010, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data", *Nucleic acids research,* vol. 38, no. 16, pp. e164.

Warthin, A.S. 1913, "Heredity with reference to carcinoma - As shown by the study of the cases examined in the Pathological Laboratory of the University of Michigan, 1895-1913", *Archives of Internal Medicine,* vol. 12, no. 5, pp. 546-555.

Webb, E.L., Rudd, M.F. & Houlston, R.S. 2006, "Colorectal cancer risk in monoallelic carriers of MYH variants", *American Journal of Human Genetics,* vol. 79, no. 4, pp. 768-71; author reply 771-772.

Win, A.K., Dowty, J.G., Cleary, S.P., Kim, H., Buchanan, D.D., Young, J.P., Clendenning, M., Rosty, C., Macinnis, R.J., Giles, G.G., Boussioutas, A., Macrae, F.A., Parry, S., Goldblatt, J., Baron, J.A., Burnett, T., Le Marchand, L., Newcomb, P.A., Haile, R.W., Hopper, J.L., Cotterchio, M., Gallinger, S., Lindor, N.M., Tucker, K.M., Winship, I.M. & Jenkins, M.A. 2014, "Risk of Colorectal Cancer for Carriers of Mutations in MUTYH, With and Without a Family History of Cancer", *Gastroenterology,* vol. 146, no. 5, pp. 1208-1211.e1-5.

Win, A.K., Macinnis, R.J., Dowty, J.G. & Jenkins, M.A. 2013, "Criteria and prediction models for mismatch repair gene mutations: a review", *Journal of medical genetics,* vol. 50, no. 12, pp. 785-793.

Yurgelun, M.B., Goel, A., Hornick, J.L., Sen, A., Turgeon, D.K., Ruffin, M.T.,4th, Marcon, N.E., Baron, J.A., Bresalier, R.S., Syngal, S., Brenner, D.E., Boland, C.R. & Stoffel, E.M. 2012, "Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps", *Cancer prevention research (Philadelphia, Pa.),* vol. 5, no. 4, pp. 574-582.

Zilonke, N., Hackl, M. & Baldaszti, E. 2014, *Krebsinzidenz und Krebsmortalität in Österreich 2014 - Herausgegeben von STATISTIK AUSTRIA,* MDH-Media, Vienna, Austria.
Zilonke, N., Hackl, M. & Baldaszti, E. 2012, *Krebsinzidenz und Krebsmortalität in Österreich 2012 - Herausgegeben von STATISTIK AUSTRIA,* MDH-Media, Vienna, Austria.

6 Appendix

nature COMMUNICATIONS

ARTICLE

Received 5 May 2014 | Accepted 8 Sep 2014 | Published 13 Oct 2014

DOI: 10.1038/ncomms6191 OPEN

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

Eduard Schulz[1], Petra Klampfl[1,2], Stefanie Holzapfel[3], Andreas R. Janecke[4,5], Peter Ulz[6], Wilfried Renner[7], Karl Kashofer[8], Satoshi Nojima[9,10], Anita Leitner[11], Armin Zebisch[1], Albert Wölfler[1], Sybille Hofer[1], Armin Gerger[12], Sigurd Lax[13], Christine Beham-Schmid[8], Verena Steinke[3], Ellen Heitzer[6], Jochen B. Geigl[6], Christian Windpassinger[6], Gerald Hoefler[8], Michael R. Speicher[6], C. Richard Boland[14], Atsushi Kumanogoh[9,15,16] & Heinz Sill[1]

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, SEMA4A^{V78M} 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.

[1] Division of Hematology, Department of Internal Medicine, Medical University of Graz, A-8036 Graz, Austria. [2] Department of Internal Medicine, General Hospital Graz West, A-8020 Graz, Austria. [3] Institute of Human Genetics, Faculty of Medicine, University of Bonn, D-53127 Bonn, Germany. [4] Division of Human Genetics, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, A-6020 Innsbruck, Austria. [5] Department of Pediatrics I, Medical University of Innsbruck, A-6020 Innsbruck, Austria. [6] Institute of Human Genetics, Medical University of Graz, A-8036 Graz, Austria. [7] Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, A-8036 Graz, Austria. [8] Institute of Pathology, Medical University of Graz, A-8036 Graz, Austria. [9] Department of Immunopathology, WPI Immunology Frontier Research Center, Osaka University, Suita City, Osaka 565-0871, Japan. [10] Department of Pathology, Osaka University Graduate School of Medicine, Osaka University, Suita City, Osaka 565-0871, Japan. [11] Department of Ear, Nose and Throat, Medical University of Graz, A-8036 Graz, Austria. [12] Division of Oncology, Department of Internal Medicine, Medical University of Graz, A-8036 Graz, Austria. [13] Department of Pathology, General Hospital Graz West, 8020 Graz, Austria. [14] Division of Gastroenterology, Baylor University Medical Center, Dallas, Texas 75246-2017, USA. [15] Department of Respiratory Medicine, Allergy and Rheumatic Disease, Graduate School of Medicine, Osaka University, Suita City, Osaka 565-0871, Japan. [16] JST, CREST, Suita City, Osaka 565-0871, Japan. Correspondence and requests for materials should be addressed to H.S. (email: heinz.sill@medunigraz.at).

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 (HNPCC) being the two major hereditary forms[2,3]. HNPCC 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 HNPCC fulfilling AC-I lack detectable germline mutations in cancer predisposition genes and are classified as familial colorectal cancer type X (FCCTX)[5–7]. In contrast to Lynch syndrome (LS)—the HNPCC 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 HNPCC syndromes[13]. 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[5], 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 Sema4A 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 Sema4A^{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].

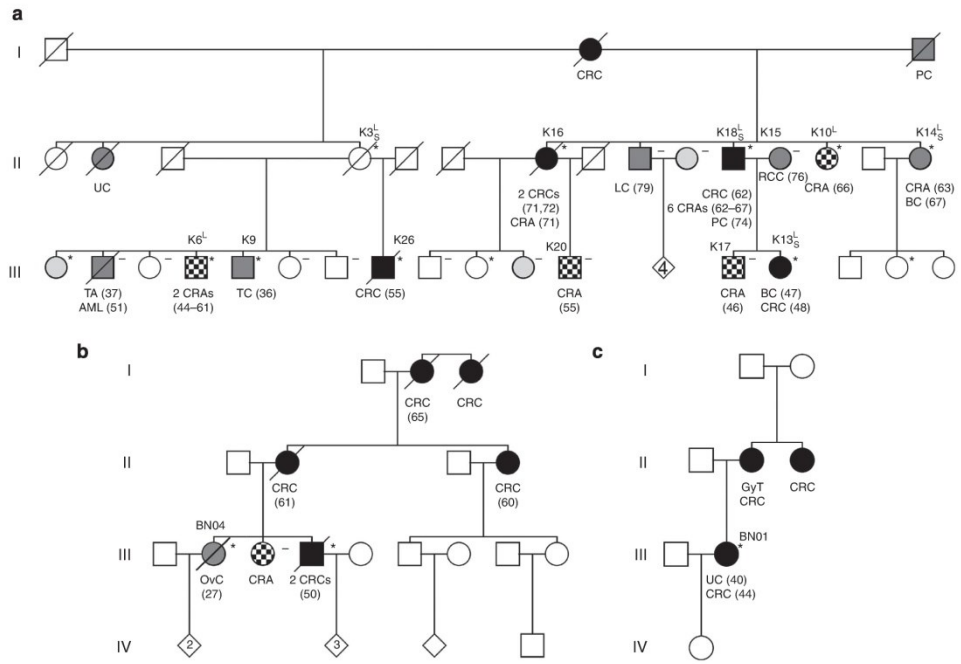


Figure 1 | Pedigrees of families with germline *SEMA4A* mutations. Families with V78M (**a**), G484A (**b**) and S326F (**c**) mutations are shown. L, individual included in LA; S, individual included in WES; asterisk, *SEMA4A* mutation carrier; minus, *SEMA4A* wild type; black symbol, CRC; checkered symbol, colorectal adenoma; dark grey, malignant neoplasm; light grey, benign neoplasm; number in symbol, number of unspecified offspring. AML, acute myeloid leukaemia; BC, breast cancer; CRA, colorectal adenoma; GyT, gynaecologic tumour; OvC, ovarian cancer; PC, prostate cancer; TA, thyroid adenoma; TC, testicular cancer; UC; uterine cancer; UT, uterine tumour. Results of mutational analyses are indicated in tested individuals only. Age at diagnosis (years) is given in parentheses. For multiple colorectal adenomas, age at first presentation or at screening colonoscopy is indicated. An extended pedigree of the family with the V78M mutation including age of the individuals is shown in Supplementary Fig. 1, histopathological characteristics of their colorectal neoplasms are summarized in Table 1.

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 SEMA4A^{V78M}-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/β-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.1451G4C, p.Gly484Ala, rs148744804; homozygous c.977C4T, 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 study[23]. 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%.

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	CRA	62	Tubular adenoma	Well to moderately differentiated	Sigmoid colon	+
K18	CRA	64	Tubular adenoma	Well to moderately differentiated	Ascending colon	+
K18	CRA	65	Tubular adenoma	Well to moderately differentiated	Descending colon	+
K18	CRA	66	Tubular adenoma	Well to moderately differentiated	NA	+
K18	CRA	67	Tubular adenoma	Well to moderately differentiated	Descending colon	+
K20	CRA	55	Tubulovillous 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.

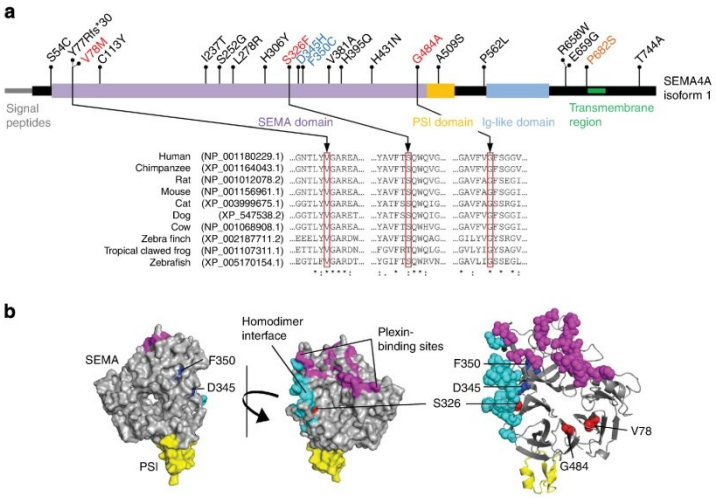


Figure 2 | Localization of germline and somatic CRC *SEMA4A* mutations at the protein level. (**a**) Germline mutations found in this study are illustrated in red, the SNP in orange, germline mutations associated with eye diseases in blue and somatic CRC mutations in black, respectively. Multiple sequence 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 cancers[24,25]. 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 growth[26]. They are characterized by an extracytoplasmic amino-terminal β-propeller—the SEMA

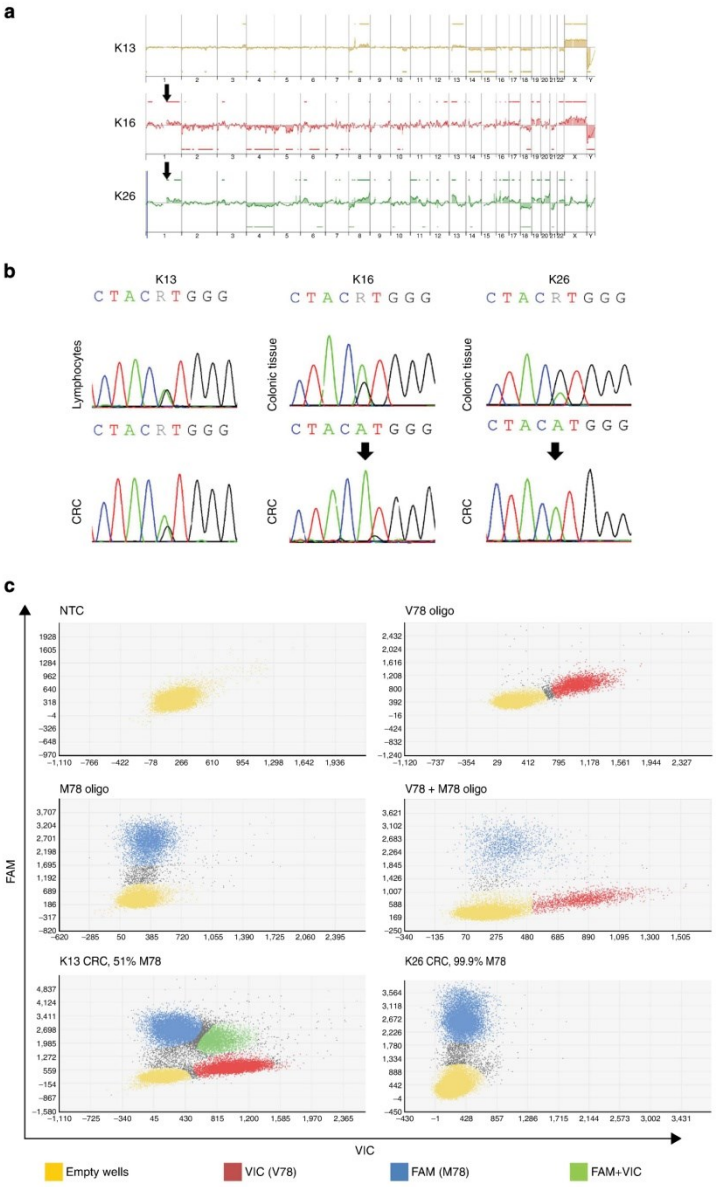


Figure 3 | LOH in CRCs of patients K16 and K26. (**a**) Array-based comparative genomic hybridization of three CRCs from Family K with germline *SEMA4A* V78M mutation. A gain in the *SEMA4A* locus is marked with an arrow. (**b**) Sanger sequencing. (**c**) Quantitative dPCR using fluorophore-coupled (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.

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_000038.5:c.2626C>T	p.R876X	rs121913333
K13	*APC* NM_000038.5:c.4348C>T	p.R1450X	rs121913332
K13	*KRAS* NM_004985.4:c.34_35delinsAT	p.G12I	NA
K13	*TP53* NM_000546.5:c.380C>A	p.S127Y	NA
K13	*PIK3CA* NM_006218.2:c.1633G>A	p.E545K	rs104886003
K16	None found	None found	—
K18	*TP53* NM_000546.5:c.844C>T	p.R282W	NA
K26	*APC* NM_000038.5:c.4135G>T	p.E1379X	rs121913326
K26	*KRAS* NM_004985.4:c.34G>A	p.G12S	NA
K26	*TP53* NM_000546.5:c.743G>A	p.R248Q	rs11540652

CRC, colorectal cancer; NA, not available.

domain—which is needed for plexin receptor binding[27,28]. In addition to their role in developmental and physiological processes, semaphorins and their receptors have increasingly been associated with neoplastic disorders (reviewed in refs 24,26). Interestingly, they have been found to act both, in an anti- and protumoral fashion depending on the particular semaphorin as well as the tumour context. Several tumorigenic properties are thereby influenced including cell proliferation, evasion of apoptosis, angiogenesis, oxidative stress regulation and metastasis. However, no particular semaphorin has been implicated in cancer susceptibility yet. Here we have shown for the first time that *SEMA4A* germline variants predispose to a hereditary neoplastic syndrome.

The germline *SEMA4A* V78M variant was inherited in an autosomal dominant fashion with incomplete penetrance in this family with FCCTX pinpointing additional genetic, environmental or life style modifiers necessary to establish the malignant phenotype. Individuals with this variant developed tumours at a higher age than classical LS patients[7], showed a moderate number of colorectal adenomas and had a propensity for extracolonic malignancies. Such a genetic modifier might be *MUTYH* where the heterozygous germline variant R217H was found in two *SEMA4A* V78M carriers with CRC (K13, K18). Biallelic germline *MUTYH* mutations—primarily Y179C and G396D—are the cause of *MUTYH*-associated polyposis, which is a rare autosomal recessive syndrome resembling familial adenomatous polyposis[12,29]. Monoallelic *MUTYH* germline variants are associated with a small increase in CRC risk; however, this assumption should be handled with care as different studies have come to inconsistent results[12,30,31]. The *MUTYH* variant R217H found in K13 and K18 has been previously described once in a cohort of 406 patients with more than five polyps and/or CRC from France but its predisposing role has not been established yet[32].

We were able to identify two further *SEMA4A* variants in a mutational screening of 53 FCCTX patients and studied the segregation of G484A, which followed a dominant inheritance pattern. In both pedigrees, variants were associated with extracolonic neoplasms—ovarian cancer in G484A and endometrial cancer in S326F. Homozygosity of the S326F genotype observed in the index patient could either be the result of an 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[2].

The compound heterozygous germline *SEMA4A* variants D345H and F350C have been described in patients with retinitis pigmentosa and cone rode dystrophy but until now this finding has not been replicated[17,33]. Sema4A-deficient mice exhibit photoreceptor degeneration and disturbed T-helper cell function but lack apparently increased tumour development[14,34]. 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 Sema4A-deficient mice in a homozygous knock-in mouse model, a genotype not described in humans yet[18]. 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[35,36]. 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[37]. 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.*[38] 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

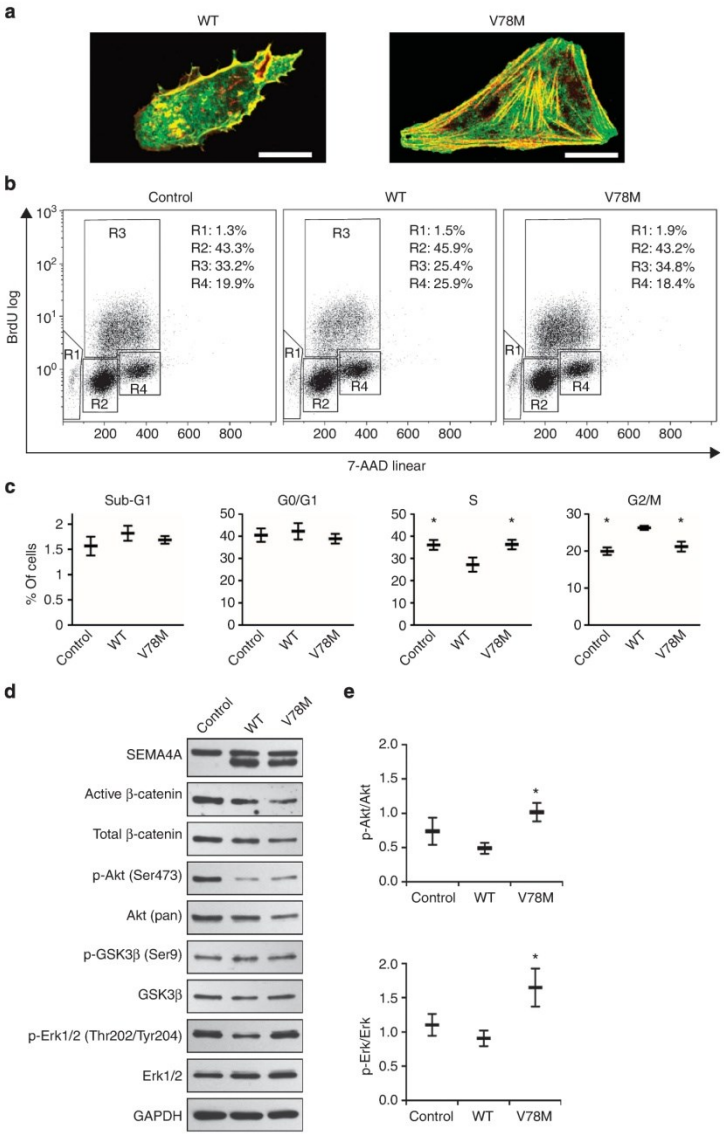


Figure 4 | SEMA4A^{V78M} shows normal surface expression and leads to cell cycle changes in HCT-116 cells. (**a**) ARPE-19 cells were transfected with the plasmid constructs expressing Sema4A^{WT}-EGFP or Sema4A^{V78M}-EGFP proteins, incubated for 48 h and stained with phalloidin. Green, Sema4A-EGFP; 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 SEMA4A^{V78M} than SEMA4A^{WT}-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) lyzed 48 h after transfection. SEMA4A^{V78M}-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.

Table 3 | Results of the *SEMA4A* Pro682Ser association study.

Cohort*	No.	Genotypes			Frequency of allele T (%)	Two-tailed *P* value†	OR (95% CI)
		CC	CT	TT			
German and Austrian FCCTX individuals	47	41	6	0	6.4	0.0008	6.793 (2.634 to 17.518)
Non-cancer controls‡	1,138	1,114	24	0	1.1	—	1

CI, confidence interval; FCCTX, familial colorectal cancer type X; OR, odds ratio.
*Genotypes of cases and controls were in Hardy–Weinberg equilibrium, with *P* values (df = 1) of 0.718 and 0.647, respectively.
†Fisher's exact test of genotype counts from cases versus controls.
‡Men, *n* = 574 (50.4%); women, *n* = 564 (49.6%); mean age: 60 years (± s.d. 18); median age: 64 years (range 15–99).

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^{V78M} 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^{V78M} 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^{V78M}.

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 criteria compatible with HNPCC (LS) were fulfilled (Fig. 1a; Supplementary Fig. 1). However, CRCs from two patients in either branch (K13 and K26) showed normal expression of the DNA MMR genes *MLH1*, *MSH2*, *MLH6* and *PMS2* by immunohistochemistry as well as microsatellite stability. Furthermore, in individual K13, tumour tissue revealed absence of a somatically acquired *BRAF* mutation. Large germline rearrangements in the MMR genes and *EPCAM*, respectively, as a rare cause of HNPCC were also excluded by multiplex ligation-dependent probe amplification in this patient. We, therefore, made a diagnosis of FCCTX in this family.

To prove that *SEMA4A* germline mutations are also operable in other patients with FCCTX, we studied a cohort of 53 further cases with this syndrome (Supplementary Table 3). Clinical data as well as DNA extracted from PB were provided by the German HNPCC Consortium, Bonn, Germany ($n = 44$), the Division of Gastroenterology, Baylor University Medical Center, Dallas, Texas, USA ($n = 6$) and the Institute of Human Genetics, MUG ($n = 3$) in accordance with local ethical guidelines. One patient (MUG1) fulfilled the modified AC criteria[4], whereas all others, classical AC-I and/or AC-II criteria.

Index patient BN01 with germline *SEMA4A* p.Ser326Phe mutation had microsatellite stable sigmoid colon cancer. Index patient BN04 identified to carry the germline *SEMA4A* p.Gly484Ala mutation has been analysed for germline mutations in *MLH1*, *MSH2*, *MLH6* and *PMS2* by direct sequencing which revealed negative results. Her brother's CRC (III:IV) showed low microsatellite instability.

DNA and RNA isolation. DNA purification from PB mononuclear cells, cell lines and fresh frozen tissue specimens were accomplished with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. The RNeasy Mini Kit (Qiagen) was used for RNA isolation from PB mononuclear cells and cell lines. To investigate tumour-specific aberrations, tumour-bearing tissue was manually microdissected from archival, FFPE specimens and DNA isolated using the ReliaPrep FFPE gDNA Miniprep kit (Promega).

Genotyping and LA. LA was performed in Family K afflicted with FCCTX (Fig. 1a; Supplementary Fig. 1). The GeneChip Human Mapping 250 K Nsp Array (Affymetrix) was used for genotyping of family members according to the manufacturer's protocols. A genome-wide analysis of linkage was conducted under the assumption of an autosomal dominant mode of inheritance with assignment of phenotype to persons affected by the trait (K3, K6, K10, K14, K13, K18) and with additional inclusion of one spouse of one affected person (K15) for improved haplotype reconstruction. The disease allele was assigned a frequency of 0.001 and 100% penetrance for multipoint parametric LA on this subset of family members which was performed with the MERLIN programme in the Alohomora Linkage software tool[43,44].

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 DNA Sample Preparation kit (Illumina), and the final products were analysed for integrity by the Agilent Bioanalyzer. Multiple DNA libraries were combined with different indices into a single pool before enrichment. Hybridization with capture probes, washing and eluting were performed two times. Enriched targeted regions were amplified by PCR using the same primers from the TruSeq DNA Sample Preparation kit and then sequenced on a HiSeq 2000 Sequencer (Illumina).

Sequence data in FastQ format were aligned to the hg19 version of the human genome (GRCh37) using the Burrows-Wheeler Aligner[45] (BWA; http://bio-bwa.sourceforge.net/), transformed into SAM files and then converted into compressed BAM files by picard (http://picard.sourceforge.net/). Possible PCR duplicates were marked by picard and local realignment around indels was performed using the Genome Analysis Tool Kit[46] (GATK; http://www.broadinstitute.org/gatk/download) to prevent false positive SNPs at the end of sequence reads. GATK was also used to reevaluate base quality scores, perform the raw SNP calling of all sequences within RefSeq gene exons (http://www.ncbi.nlm.nih.gov/RefSeq/) - plus ten bp at each splice site—and to recalibrate variant quality scores.

With a read length of 101 bp, there were, on average, 88,333,643 total reads that could be mapped to the human genome in 64.5%, respectively. The mean read depth of target regions (96.4% of RefSeq (refGene) coding exons and 97.2% of CCDS coding exons, respectively) was 49.1 × . The mean coverage of target regions more than 1 × was 94.5% and the mean coverage of target regions more than 10 × was 87.3%, respectively.

Variant calls were annotated with ANNOVAR[47] (http://www.openbioinformatics.org/annovar/), which contained the data from dbSNP132 (http://www.ncbi.nlm.nih.gov/SNP/) and the allele frequencies of the 1000 Genomes Project from February 2012 (http://www.1000genomes.org/) and of the ESP5400 version of the NHLBI GO Exome Sequencing Project (https://esp.gs.washington.edu/drupal/). During progression of the study, variants were also manually checked for frequencies in updated versions of those databases. Furthermore, single variants were analysed by the following prediction programs: SIFT (http://sift.jcvi.org/), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), MutationTaster 2.0 (http://www.mutationtaster.org/), PhyloP, phastCons and GERP (the last three were precalculated from the UCSC genome browser http://genome.ucsc.edu/cgi-bin/hgGateway). An in-house databank consisting of 18 exomes sequenced on the Illumina platform was used to exclude sequence artifacts as well as variants not covered extensively by public databases. The median age of

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 (Qiagen) or the peqGOLD 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 v.1.4.0 (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 *B2M* were as previously described[48]. Primers for *SEMA4A* transcript variants were as follows: var1-3fw, 5′-CCTGGGCCTTTTCCTCTTCC-3′; var124fw, 5′-TTTCTCCTGAATGGCAC CCC-3′; var1-4rv, 5′-TTTTTCTGTCACTGGCTGGC-3′ (the reverse primer was the same for all transcript variants). Primers var1-3fw and var1-4rv 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[49]. 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 2,855 COSMIC annotated hot spot mutations, were amplified by multiplexed PCR using the IonAmpliSeq Cancer Hotspot 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 (Ion 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 fragment size-range. On average, one million reads were obtained for each sample with more than 90% of bases above AQ20 and 87 to 93% reads on-target. Sequence information was obtained from tumour samples in duplicates and additionally from normal non-tumour material.

Initial data analysis was performed using the Ion Torrent Suite Software (Thermo Fisher Scientific, open source, GPL, https://github.com/iontorrent/). Briefly, this included base calling, alignment to the reference genome (hg19) using the TMAP mapper and variant calling with a modified diBayes approach taking into account the flow space information. All called variants were annotated using open source software[47,50] (ANNOVAR, http://www.openbioinformatics.org/annovar/; SnpEff, http://snpeff.sourceforge.net/) and custom Perl scripts. Coding, non-synonymous sequence variations that were detected and confirmed in tumours but not in the normal tissue were further evaluated and visually inspected in IGV (http://www.broadinstitute.org/igv/) to exclude erroneous variant calls resulting from PCR artifacts or sequence effects. The detection threshold was set to 10% mutated alleles in both duplicates.

Array comparative genomic hybridization. Tumour DNA samples were labelled using the BioPrime Array CGH Genomic Labeling System (Invitrogen) according to manufacturer's protocol. Briefly, 250 ng of AluI and RsaI digested tumour and reference DNA (Promega) were differentially labelled with dCTP-Cy5 and dCTP-Cy3, respectively (GE Healthcare) and purified by Amicon Ultra-0.5 30kDA filters (Millipore). Analysis of DNA copy-number changes was conducted using a SurePrint G3 60K array (Agilent) scanned on the microarray scanner G2505B (Agilent) according to the manufacturer's instructions. Feature Extraction and DNA Workbench softwares (Agilent) were used for data analysis.

Digital PCR. The *SEMA4A* V78M mutation was quantitatively analysed with digital PCR (dPCR) on the QuantStudio 3D platform (Life Technologies). A Custom TaqMan SNP Genotyping Assay specific for the analysis of the V78M mutation was used and tested on a StepOne Plus instrument (Life Technologies) using the TaqMan Genotyping Master Mix (Life Technologies) according to the manufacturer's recommendations. For dPCR, 17.4 μl of Digital PCR Master Mix (2 ×) was mixed with 1.7 μl of the TaqMan assay and 60 ng of DNA to a final volume of 36 μl and subjected to two Digital PCR 20k Chips. The chips were thermally cycled in a two-step PCR using the GeneAmp PCR System 9700 (10 min 96 °C, followed by 44 cycles of 56 °C 2 min and 94 °C 30 s, final extension of 2 min 58 °C) and imaged in the QuantStudio 3D instrument. Raw data were analysed using the Relative Quantification module of the QuantStudio 3D AnalysisSuite Software. The confidence level was set to 95% and the desired precision value was 10%.

Cell culture. Adherent cell lines HT-29, SW-480, HCT-116, HRT-18 and 293T were obtained from ATCC and cultivated for a maximum of 6 weeks in DMEM (Sigma-Aldrich) supplemented with 10% (v/v) HyClone fetal bovine serum (Thermo Scientific) and 1X Antibiotic-Antimycotic (Life Technologies) in a humidified chamber at 37 °C and 5% CO_2. The identity of all cell lines was confirmed by VNTR analysis using the AmpF/STR Profiler Plus Kit and ABI PRISM 310 Genetic Analyzer (both by Applied Biosystems, respectively) according to the manufacturer's protocols and verified at the online service of the DSMZ cell bank (http://www.dsmz.de).

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, 6×10^5 cells were seeded into six-well tissue culture plates to achieve 60 to 80% confluency. Plasmid and Lipofectamine LTX (Life Technologies) were diluted at a ratio of 1:10 (HCT-116) or 1:5 (293T), respectively, in 500 μl serum-free 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 incubating on a shaker 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 SPECTROstar Omega and MARS Data Analysis Software (both BMG LABTECH). Lysates were diluted with 4 × Laemmli sample buffer (Bio-Rad) and 710 mM 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 (Advansta) or Supported Nitrocellulose (Bio-Rad) membranes, respectively. Membranes were blocked with 3% (wt/v) 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) (1:1,500, #4691, Cell Signaling), Phospho-Akt (Ser473) (1:2,000, #4060, Cell Signaling), p44/42 MAPK (Erk1/2) (1:2,500, #4695, Cell Signaling), Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (1:2,000, #4370, Cell Signaling), GSK-3β (1:1,500, #9832, Cell Signaling), Phospho-GSK-3β (Ser9) (1:3,000, #5558, Cell Signaling), GAPDH (1:2,000, #sc-32233, Santa Cruz Biotechnology). Horseradish peroxidase-linked secondary antibodies were anti-rabbit IgG (#7074, Cell Signaling) and anti-mouse immunoglobulins (#P026002, 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 (Advansta), CL-XPosure films (Thermo Scientific) and CURIX 60 developer (Agfa Healtcare), respectively. ImageJ 1.47v (NIH, rsbweb.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 described[18]. 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 Sema4AV78M-EGFP construct was generated from Sema4AWT-EGFP using a QuikChange 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 Biolabs) 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 monitored for 48 h with a 1-h interval by the Cell Observer (Carl Zeiss). ImageJ 1.47v 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, 1.5×10^6 HCT-116 cells were transferred 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 washed with ice cold PBS by centrifugation at 4 °C and then fixed for 30 min at room temperature, permeabilized for 10 min on ice, refixed for 5 min at room temperature, treated with DNase and finally stained with APC anti-BrdU antibody (1:50 for 30 min at room temperature) as well as 7-AAD according to the instructions from the APC BrdU Flow Kit (BD Pharmingen). Unlabelled native cells were used as a negative control for the APC anti-BrdU antibody. Stained cells were acquired on the BD LSR II Flow Cytometer operated with FACSDiva Software (both from BD Biosciences, respectively) with a flow rate of less than 400 cells s^{-1} on the same day of staining. Kaluza Flow Cytometry Analysis Software v1.2 (Beckman Coulter) was used for analysis and illustration of flow cytometry data.

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 POLYVIEW-3D (http://polyview.cchmc.org/polyview3d.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. Jasperson, 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-I 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,** 422–428 (2012).
11. Sill, H., Olipitz, W., Zebisch, A., Schulz, E. & Wolfler, A. Therapy-related myeloid neoplasms: pathobiology and clinical characteristics. *Br. J. Pharmacol.* **162,** 792–805 (2011).
12. Farrington, S. M. *et al.* Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am. J. Hum. Genet.* **77,** 112–119 (2005).
13. Yurgelun, M. B. *et al.* Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps. *Cancer Prev. Res. (Phila)* **5,** 574–582 (2012).
14. Kumanogoh, A. *et al.* Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice. *Immunity* **22,** 305–316 (2005).
15. Delgoffe, G. M. *et al.* Stability and function of regulatory T cells is maintained by a neuropilin-1-semaphorin-4a axis. *Nature* **501,** 252–256 (2013).
16. Toyofuku, T. *et al.* Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1. *EMBO J.* **26,** 1373–1384 (2007).
17. Abid, A., Ismail, M., Mehdi, S. Q. & Khaliq, S. Identification of novel mutations in the SEMA4A gene associated with retinal degenerative diseases. *J. Med. Genet.* **43,** 378–381 (2006).
18. Nojima, S. *et al.* A point mutation in Semaphorin 4A associates with defective endosomal sorting and causes retinal degeneration. *Nat. Commun.* **4,** 1406 (2013).
19. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
20. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36,** 320–328 (2011).
21. Shtivelman, E., Sussman, J. & Stokoe, D. A role for PI 3-kinase and PKB activity in the G2/M phase of the cell cycle. *Curr. Biol.* **12,** 919–924 (2002).
22. Shinohara, M., Mikhailov, A. V., Aguirre-Ghiso, J. A. & Rieder, C. L. Extracellular signal-regulated kinase 1/2 activity is not required in mammalian cells during late G2 for timely entry into or exit from mitosis. *Mol. Biol. Cell* **17,** 5227–5240 (2006).
23. Bahadori, B. *et al.* Polymorphisms of the hypoxia-inducible factor 1 gene and peripheral artery disease. *Vasc. Med.* **15,** 371–374 (2010).
24. Rehman, M. & Tamagnone, L. Semaphorins in cancer: biological mechanisms and therapeutic approaches. *Semin. Cell Dev. Biol.* **24,** 179–189 (2013).
25. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).
26. Neufeld, G., Sabag, A. D., Rabinovicz, N. & Kessler, O. Semaphorins in angiogenesis and tumor progression. *Cold Spring Harb. Perspect. Med.* **2,** a006718 (2012).
27. Janssen, B. J. *et al.* Structural basis of semaphorin-plexin signalling. *Nature* **467,** 1118–1122 (2010).
28. Nogi, T. *et al.* Structural basis for semaphorin signalling through the plexin receptor. *Nature* **467,** 1123–1127 (2010).
29. Croitoru, M. E. *et al.* Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J. Natl Cancer Inst.* **96,** 1631–1634 (2004).
30. Win, A. K. *et al.* Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. *Gastroenterology* **146,** 1208–11.e1-5 (2014).
31. Webb, E. L., Rudd, M. F. & Houlston, R. S. Colorectal cancer risk in monoallelic carriers of MYH variants. *Am. J. Hum. Genet.* **79,** 768–771 author reply 771-2 (2006).
32. Olschwang, S., Blanche, H., de Moncuit, C. & Thomas, G. Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the MYH gene identified in a population with adenomatous polyposis. *Genet. Test.* **11,** 315–320 (2007).
33. Ismail, M., Abid, A., Anwar, K., Mehdi, S. Q. & Khaliq, S. Refinement of the locus for autosomal recessive cone-rod dystrophy (CORD8) linked to chromosome 1q23-q24 in a Pakistani family and exclusion of candidate genes. *J. Hum. Genet.* **51,** 827–831 (2006).
34. Rice, D. S. *et al.* Severe retinal degeneration associated with disruption of semaphorin 4A. *Invest. Ophthalmol. Vis. Sci.* **45,** 2767–2777 (2004).
35. Baker, S. M. *et al.* Enhanced intestinal adenomatous polyp formation in Pms2 − / − ;Min mice. *Cancer Res.* **58,** 1087–1089 (1998).
36. Takaku, K. *et al.* Intestinal tumorigenesis in compound mutant mice of both Dpc4 (Smad4) and Apc genes. *Cell* **92,** 645–656 (1998).
37. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339,** 1546–1558 (2013).
38. Middeldorp, A. *et al.* Increased frequency of 20q gain and copy-neutral loss of heterozygosity in mismatch repair proficient familial colorectal carcinomas. *Int. J. Cancer* **130,** 837–846 (2012).
39. Luchino, J. *et al.* Semaphorin 3E suppresses tumor cell death triggered by the plexin D1 dependence receptor in metastatic breast cancers. *Cancer Cell* **24,** 673–685 (2013).
40. Sanchez-de-Abajo, A. *et al.* 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 (2007).
41. Francisco, I. *et al.* Familial colorectal cancer type X syndrome: two distinct molecular entities? *Fam. Cancer* **10,** 623–631 (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, I. 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, Leukämiehilfe 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 PhD 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., C.B.-S., V.S., 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., A.K. 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).