Received: 15 April 2019 | Revised: 26 June 2019 | Accepted: 27 June 2019

DOI: 10.1002/gcc.22786

RESEARCH ARTICLE

WILEY

Genetic analyses supporting colorectal, gastric, and prostate cancer syndromes

Karin Wallander[1] | Wen Liu[2] | Susanna von Holst[2] | Jessada Thutkawkorapin[2] | Vinaykumar Kontham[2] | Anna Forsberg[3] | Annika Lindblom[1] | Kristina Lagerstedt-Robinson[1]

[1]Department of Molecular Medicine and Surgery, Karolinska Institutet, and Department of Clinical Genetics, Karolinska University Hospital, Solna, Stockholm, Sweden

[2]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

[3]Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Correspondence
Annika Lindblom, Department of Clinical Genetics L5:03, Karolinska University Hospital, Solna, SE-17176 Stockholm, Sweden.
Email: annika.lindblom@ki.se

Funding information
Cancerfonden, Grant/Award Number: 160458; National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services, Grant/Award Number: U19 CA148107 R01 CA81488 P30 CA014089 R01 CA197; Stockholm County Council, Grant/Award Number: ALF; The Stockholm Cancer Society, Grant/Award Number: 161183; Vetenskapsrådet, Grant/Award Number: K2015-55X-22674-01-4 K2008-55X-20157-03-3; Swedish Research Council; Science for Life Laboratory; Akademiska Sjukhuset; Uppsala Universitet

Abstract

Colorectal cancer (CRC), prostate cancer (PrC), and gastric cancer (GC) are common worldwide, and the incidence is to a certain extent dependent on genetics. We have recently shown that in families with more than one case of CRC, the risk of other malignancies is increased. We therefore suggested the presence of not yet described CRC syndromes. In this study, we have searched for genetic susceptibility loci for potential cancer syndromes involving CRC combined with PrC and/or GC. We have performed SNP (single-nucleotide polymorphism)-based linkage analyses in 45 families with CRC, PrC, and GC. In the regions with suggested linkage, we performed exome and association haplotype analyses. Five loci generated a high logarithm of odds (HLOD) score >2, suggestive of linkage, in chromosome bands 1q31-32, 1q24-25, 6q25-26, 18p11-q11, and Xp11. Exome analysis detected no potential pathogenic sequence variants. The haplotype association study showed that one of the top five haplotypes with the lowest *P* value in the chromosome band 6q25 interestingly was found in the family which contributed the most to the increased HLOD at that locus. This study supports a suggested hereditary cancer syndrome involving CRC and PrC and indicates a location at 6q25. The impact of this locus needs to be confirmed in additional studies.

KEYWORDS
colon, gastric cancer syndromes, prostate

1 | INTRODUCTION

Heritability has been estimated to account for 35% of the variation in susceptibility of colorectal cancer (CRC),[1] and in less than 5% of the cases, the genetic cause is a known cancer syndrome such as Lynch syndrome (caused by mutations in the genes *MLH1*, *MSH2 [EPCAM]*, *MSH6*, or *PMS2*) or familial adenomatous polyposis (mutations in the *APC* or the *MUTYH* genes).[2] Rare CRC syndromes and low penetrance genes are thought to explain some of the remainder of the hereditary CRC cases, but the vast majority is still unknown.[3]

CRC is generally considered to develop through the adenoma-carcinoma pathway.[4,5] Today, three main mechanisms defined by the underlying molecular pathology have been suggested: the chromosomal instability pathway, the microsatellite instability pathway, and the epigenetic pathway.[4,6,7] An overlap between these pathways within one tumor is likely.[8] Adenoma is considered a precursor lesion of CRC, and

Karin Wallander and Wen Liu contributed equally to this study.

some adenomas are believed to be more prone to become malignant than others.[7,9] It seems like most, but not all, adenomas have the potential of converting into malignant lesions, at least in the colon and rectum.[10] The adenoma incidence is increased in some families with prostate cancer (PrC) and gastric cancer (GC), even when no known syndrome is yet diagnosed, which suggests a shared genetic etiology in these families.[11] It has also been demonstrated that the risk of colorectal adenomas is increased in GC patients.[12,13]

Recently, we have shown that the risk of malignancies such as GC and PrC is increased in families with familial CRC.[14] The aim of this study was to further examine the suggested new syndromes using genetic analyses in families with these tumor types and see if we could define potential new genetic susceptibility loci. We performed linkage analyses in families with CRC and high-risk adenomas combined with PrC and/or GC. To increase the number of informative family members, we also did a separate additional linkage analysis, where individuals with all subtypes of adenoma, regardless of size and degree of dysplasia, were considered to be affected. We then conducted whole exome sequencing (WES) in the families showing linkage to the suggested loci to look for clinically hazardous potential sequence variants in the coding genome. To further investigate the loci showing linkage, we executed an association haplotype study in families with CRC.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

The study was undertaken in accordance with the Swedish legislation of ethical permission (2003:460) and according to the decision in the Stockholm regional ethical committee (ref 2002/489, 2003/198, and 2008/125-31.2).

2.2 | Individuals in the linkage study, sequencing analyses, and association haplotype study

Families included in the linkage study and WES were recruited among patients referred to the Department of Clinical Genetics at Karolinska University Hospital in Stockholm, Sweden, between 1990 and 2005. Lynch syndrome was excluded using a standardized clinical protocol,[15] and polyposis syndromes were excluded using medical records. The inclusion criteria for a family were at least two individuals with CRC and at least one individual with PrC and/or GC. No other patient selection criteria were used. A family was defined as related individuals within three generations. In total, 45 families (named families 8, 12, 26, 91, 141, 177, 185, 210, 227, 229, 288, 296, 310, 324, 348, 350, 445, 485, 535, 578, 588, 611, 644, 761, 778, 779, 794, 798, 815, 836, 849, 871, 897, 918, 975, 1042, 1075, 1117, 1123, 1164, 1193, 1218, 1290, 1298, and 1300) containing 211 genotyped individuals were included in the linkage study (Table S1). All families were assigned to Nordic descent according to family name.

In the association haplotype study CRC cases were recruited in a multicenter study, The Swedish Low-risk Colorectal Cancer Study, and included all consecutive new CRC patients from 14 surgical clinics in Sweden between 2004 and 2009.[14] One indivdual from a healthy twin pair from the Swedish Twin registry was used as normal controls. A total of 685 CRC, GC, and PrC syndrome cases and 4780 controls were used in the haplotype analysis. Another cohort consisting of 54 CRC families was used (including 10 families from the initial linkage analysis cohort) to search for suggested risk haplotypes among other families. One affected patient and one first-degree relative from each family were analyzed.

2.3 | Genotyping in the linkage analysis

In the linkage analysis, the genetic analyses were performed as described,[16] although this time 45 families with at least one case of PrC or GC were used and subdivided into three different analysis groups: all 45 families were analyzed in one CRC, PrC, and GC syndrome group, 32 families in the CRC and PrC syndrome groups, and 22 families in the CRC and GC syndrome groups (see Table 1). The analysis was executed twice for all three groups. Firstly, only individuals with CRC and advanced adenomas (showing high degree of dysplasia) were coded as affected. Secondly, individuals with CRC and any adenomas were coded as affected. By this approach, 18 additional individuals could be included as affected in the second analysis. Merlin by default allows a maximum of 24 bits for each family, and therefore one large family, family 26, had to be split into three families. The family was divided so that each subfamily used one common ancestor and fitted into the limit as defined while running the program. Therefore, 45 families were analyzed as 47 families. A high logarithm of odds (HLOD) score >2 was considered suggestive of linkage.

2.4 | Whole exome sequencing

In the regions with suggestive linkage, we analyzed exome sequencing data from single individuals in families contributing to the increased HLOD score. WES was performed at Science for Life Lab, Stockholm. Briefly, pair-end sequencing samples were prepared according to the manufacturer's instructions (TrueSeq, Illumina; Sure Select, Agilent) and sequenced on an Illumina HiSeq 2000 instrument. After sequencing, the reads were aligned to the reference genome hg19/GRCh37 using Burrows-Wheeler Aligner (BWA).[17] The reads were sorted and PCR duplicates were removed with Picard. Calculation of mapping and enrichment statistics were performed with Picard[18] and genome analysis toolkit (GATK). Variants were called using GATK and followed a best practice procedure implemented at the Broad Institute.[19] The output sequence variants were annotated using ANNOVAR.[20]

Variants were filtered in several steps. Only variants in the exons or on splice sites were included. Synonymous variants were excluded as well as variants with a reported population frequency above 20%.[21-23] Data from a local population CRC database were used to assess the frequency of each specific variant. A ratio between the local CRC variant frequency and the population frequency was calculated, and variants with a ratio above 1.5 were included. Sequence variants in the analyzed loci occurring in all sequenced and affected individuals from the families showing a LOD score >1 (family 918 for the loci in chromosome bands 1q31-32, 1q24-25, 18p11-q11, and Xp11 and family 26 for the locus in chromosome band 6q25-26) were included. This is because a LOD score >1 in one individual family was considered substantially

TABLE 1 Demographic data in the linkage analysis and number of families included in the linkage analysis and their diagnoses sorted by cancer syndrome subgroup

Cancer syndrome group[a]	Number of families	Number of genotyped individuals	Number of individuals with CRC	Number of individuals with GC	Number of individuals with PrC
CRC, PrC, GC	45	439	146	36	45
CRC, PrC	31	321	104	11	45
CRC, GC	23	227	73	35	11

Abbreviations: CRC, colorectal cancer; GC, gastric cancer; PrC, prostate cancer.
[a]Regardless of adenoma classification.

contributing to the overall HLOD. Finally, if the inheritance model, originating from the results of the linkage analysis, was autosomal recessive, only genes with two or more remaining sequencing variants or variants occurring in a homozygous state were included. If the model was dominant, the variant was only considered if at least one second family contributing, though not substantially, to the HLOD score at the locus had a filtered sequence variant in the same gene.

2.5 | Association haplotype study

In the association haplotype study, DNA was extracted from peripheral blood samples for both cases and controls. The cases and controls were selected from a larger cohort, originally consisting of 2690 cases and 4782 controls. The cases were genotyped at the Center for Inherited Disease Research at Johns Hopkins University, USA, using the Illumina Infinium OncoArray-500 K BeadChips. The controls from the Swedish Twin registry were genotyped in Uppsala, Sweden, using the Illumina OmniExpress bead chip or the Illumina Infinium PsychArray-24 BeadChip. All samples went through quality control (QC) at their corresponding centers, and in total 240 370 SNPs (single nucleotide polymorphisms) were shared between the two platforms and could be used for analysis. The data were merged and the TOP strand format was accounted for. In total, 7472 individuals were proceeded for additional QC analysis.

In the first QC round (QC1), heterozygous haploid genotypes were excluded as well as samples with gender inconsistency and same position variants. 239 113 SNPs and 7472 individuals (2690 cases and 4782 controls) passed QC1. A second QC stage (QC2) was performed on the merged data, where SNPs with <98% call rate, <1% minor allele frequency, and those inconsistent with Hardy-Weinberg equilibrium (HWE 0.001) in controls were removed. 224 210 SNPs remained after QC2. In the third and final QC (QC3), a multidimensional scaling (MDS) analysis was conducted on all the remaining markers for the purpose of population stratification and to identify ethnic outliers. These outliers were excluded from the data set while the remaining were plotted in an MDS plot. After QC3, 219 114 SNPs and 7417 individuals (2637 cases, 4780 controls) remained.

PLINK V1.07[24] was used in the association haplotype studies for all five loci suggested by linkage analysis. The association study was performed with sliding window sizes up to 30. Bonferroni-adjusted *P* value criteria for genome-wide statistical significance of SNP was calculated for each locus by dividing 0.05 by the number of tests.

2.6 | Genotyping of familial samples to test for haplotypes

Genotyping was performed by the Illumina Infinium assay using the Illumina HumanOmniExpress-12v1_H BeadChip. The results on 730 525 SNPs were analyzed using the software GenomeStudio 2011.1 from Illumina Inc. The average sample call rate per SNP with sample call rate above zero was >99%, and the overall reproducibility was >99.99%. Arrays were processed according to manufacturers' protocol at the SNP&SEQ Technology Platform at Uppsala University and is available on request.[25]

3 | RESULTS

3.1 | Linkage analysis

Linkage analysis was performed to look for loci connected to a putative novel cancer syndrome involving CRC and GC and/or PrC.

In the first analysis, where CRC, advanced adenomas, PrC, and GC were coded as affected, one locus showed an HLOD score >2. As can be seen in Table 2, the region spanned approximately 6.3 Mb in chromosome bands 1q31.3-32.1, with the boundary limited to markers in each direction not showing an HLOD score >2.

In the following linkage analyses, where individuals with small adenomas were coded as affected, five loci generated an HLOD scores >2, in chromosome bands 1q31-32, 1q24-25, 6q25-26, 18p11-q11, and Xp11. There were two different HLOD scores close to three, in chromosome bands 1q24-25 and Xp11, both occurring in the CRC, PrC, and GC syndrome group assuming a recessive inheritance model. The locus in chromosome bands 1q31-32 showing linkage in the first analysis also showed linkage in the same location in the second analysis.

Family 918 and family 26 were the only families with LOD scores >1 in any of the loci that showed linkage. Family 918 was the main contributor to the increased total HLOD score at multiple positions and the family showing the highest individual LOD score (1.8, occurring in the same locus as the overall maximal HLOD score on chromosome 1q31.3-32.1). Figure 1 shows the pedigrees of family 918 and family 26.

3.2 | Whole exome sequencing

We examined the five regions that showed the suggestive linkage to the CRC syndromes for potential pathogenic sequence variants using

TABLE 2 All chromosomal loci with a HLOD score >2

Coding criteria	Syndrome group	Inheritance pattern	HLOD	Chromosome	Region interval		Nucleotide interval (basepair position)		Length (Mbp)	Family with LOD score >1
Advanced adenomas[a]	CRC, GC	AR	2.11	1q31.3-32.1	rs149067	rs1325309	194824755	201079458	6.3	918
All adenomas[b]	CRC, GC	AR	2.12	1q31.3–32.1	rs149067	rs1325309	194824755	201079458	6.3	918
All adenomas[b]	CRC, GC, PrC	AR	2.87	1q24.2-25.3	rs2902569	rs1325747	168117866	182337674	14.2	918
All adenomas[b]	CRC, GC, PrC	AD	2.28	6q25.3-26	rs1832871	rs1333962	158722034	163268904	4.5	26
All adenomas[b]	CRC, PrC	AD	2.45	6q25.2-26	rs686761	rs1333962	153209521	163268904	10.1	26
All adenomas[b]	CRC, GC, PrC	AR	2.32	18p11.2-q11.2	rs1005930	rs1972602	10143877	22113964	12.0	918
All adenomas[b]	CRC, GC, PrC	AR	2.98	Xp11.4-11.23	rs206037	rs2187789	39493720	46836070	7.3	918

Note: The chromosomal regions that showed a suggested linkage in the linkage analysis and the syndrome groups they occurred in. All positions are annotated according to GRCh37.

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; CRC, colorectal cancer; GC, gastric cancer; PrC, prostate cancer.

[a]Family members with CRC, GC, PrC, and advanced colorectal adenomas were coded as affected.

[b]Family members with CRC, GC, PrC, and all types of adenomas were coded as affected.

WES analysis. No sequence variants fulfilled the selection criteria on any of the loci.

3.3 | Association haplotype study

As no pathogenic sequence variant was found in any of the loci with suggested linkage to a CRC syndrome, haplotype association studies were also performed for each of the five loci. No locus with a significant *P* value was found. The five haplotypes with the lowest *P* value and odds ratio >1 at each locus are shown in Table 3. Those haplotypes were further searched for in the familial CRC cohort. None of the SNPs in the haplotype on chromosome X was genotyped in the cohort, and therefore this haplotype could not be studied. For all other loci, several families could potentially have the suggested haplotype, although not all markers were informative for any haplotype. Family 26 (Figure 1A), which contributed the most to the HLOD score in chromosome bands 6q25-26 in the linkage analysis, showed an almost perfect match for one of the haplotypes with the lowest *P* value (.000036) within that locus from the association haplotype study (Figure 2). Family 918 (Figure 1B), which contributed the most to the HLOD score on all loci except that in chromosome bands 6q25-26 in the linkage analysis, did not carry any of the suggested haplotypes.

4 | DISCUSSION

Cancer susceptibility loci are generally sought in families with an isolated cancer diagnosis, not in combination with other cancers. So far, we do not know of any described monogenetic cancer syndrome including highly penetrant CRC, PrC, and/or GC. We have recently shown that hereditary CRC often seems to be associated with an increased risk of malignancies at other sites. Within a cohort of patients with familial CRC, a significantly increased risk of other malignancies was observed, among them were GC, PrC, urinary bladder cancer, and malignant melanoma.[14] In the present study, we focus only on the presence of GC and PrC because we were able to collect a large enough cohort of these patients with verified diagnoses. Therefore, the aim of this study was to search for genetic loci linked to these potential cancer syndromes.

As it is unclear whether there is one syndrome with both CRC, PrC, and GC or syndromes with CRC and either PrC or GC, we analyzed the material in three ways. First using all 45 families, next the families with cases of CRC and PrC with or without GC, and finally the families with cases of CRC and GC with or without PrC. We also tested two approaches: one coding only individuals with advanced colorectal adenomas as affected and one less strict, where all individuals with adenomas were coded as affected. None of the linkage analyses resulted in a statistically significant HLOD score (>3). We decided to further study loci with an HLOD >2.

Patients with adenomas at clinical examination may develop cancer later in life and thereby they could increase the statistical power of the linkage analysis if coded as affected. Approximately 50% of adenomas with a size more than 20 mm develop into CRC. The corresponding

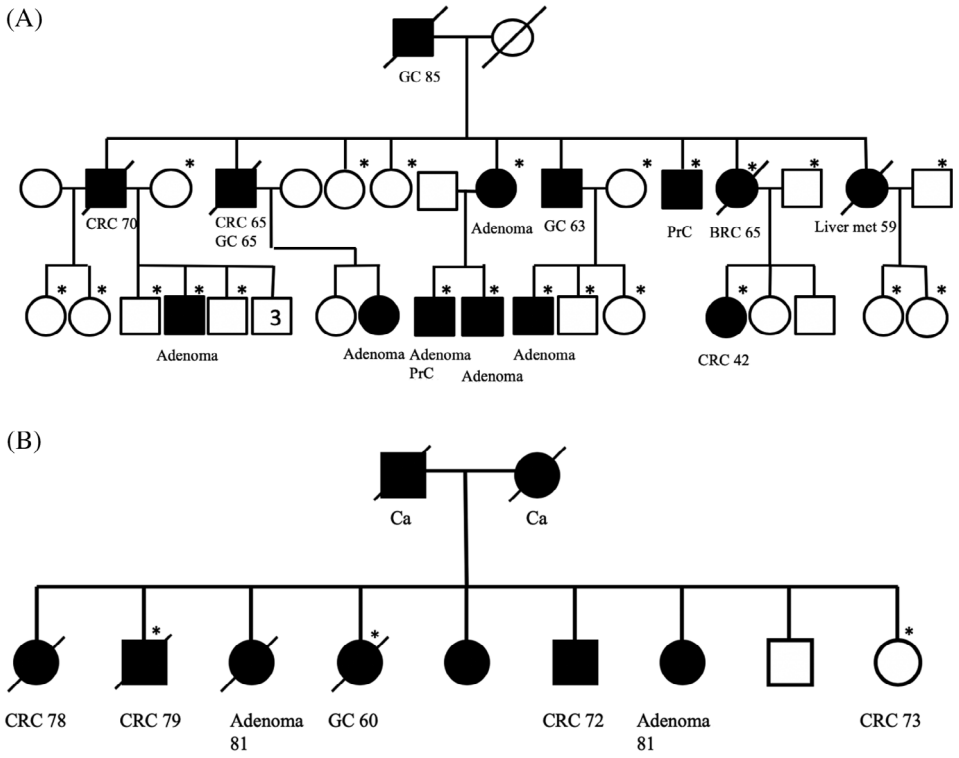

FIGURE 1 Pedigrees of the families contributing the most to the increased HLOD score. Pedigrees of family 26 (A) and family 918 (B), which contributed the most to the HLOD score on the loci with a suggestive linkage in the linkage analysis

figure in adenomas <1 cm is about 1%. Small adenomas are clearly not as prone to become malignant as large adenomas, but nevertheless, some of them will develop into CRC according to the adenoma-carcinoma pathway principle.[10] Coding all adenomas as affected could increase the statistical power of the linkage analysis and might reveal more CRC susceptibility loci by increasing the number of informative individuals. In the second linkage analysis, coding all adenomas as CRC, an HLOD score >3 was not seen but five loci had an HLOD score >2, suggestive of linkage. No confirmation of the already known or suggested CRC, PrC, or GC loci was possible in the present study, thus this study reflects potential novel loci for these suggested cancer syndromes.

The family contributing the most to the increased HLOD score in general, family 918, showed the same maximal LOD score, regardless of whether individuals with adenomas were coded as affected or unaffected, because no individual with adenoma was genotyped in that family. In family 26, which was the family with a major contribution to the increased HLOD score in chromosome bands 6q25-26, the LOD score was increased from zero to >1 when all adenomas were coded as affected, because more individuals were informative in the statistical analysis.

In the region in chromosome bands 6q25-26, the syndrome group including PrC and not GC showed a higher HLOD score, and it spanned a longer distance than when GC was included. This indicates a higher likelihood of an isolated CRC-PrC syndrome.

A disadvantage in many linkage studies is the size of the families included and the number of cases affected. In comparison to other linkage analysis reports, this study cohort consists of a large number of families, but unfortunately not all persons with a cancer diagnosis could be genotyped. The size of the families is a factor affecting linkage results; the smaller the families the lower the HLOD score in general. In order for the HLOD score to show linkage, there must be multiple families with linkage to the same region and the families in this study might very well have disease-causing variants in different loci, resulting in the lack of multiple accompanying LOD scores for these susceptibility loci. An even more comprehensive study could generate a higher HLOD score. The reason no locus showed an HLOD score >3 in our study might obviously also be that there is no linkage to a specific locus in the hypothesized syndrome groups involving CRC, PrC, and/or GC. The patients in this linkage study were included because there were at least two CRC cases in the family, which poses a bias for phenotypes with a high penetrance for CRC specifically and not the other cancer forms. Another bias in the study is the fact that only patients referred for genetic counseling were included.

To investigate the loci that showed linkage, we analyzed exome data from families contributing the most to the increased HLOD score. We searched for a gene variant that segregated within the families. No potential high-risk sequence variant was found.

It is important to keep in mind that exome analyses only call genetic variants in the coding part of the genome and large deletions or duplications cannot easily be detected. Inter- and intragenic sequence variants, controlling splicing, expression of genes, and so on, might be the underlying reason for the suggested syndromes, and they would not be discovered using exome analysis.

As no apparent pathogenic variants in the coding part of the genome could be found in the loci showing linkage to the suggested cancer syndromes, we proceeded with a targeted association haplotype study of the loci. Then, in a follow-up analysis, the top five haplotypes with the lowest *P* value and odds ratio >1 in each locus were searched for in a second population including families used in the linkage analysis. All the haplotypes are rare in the affected population, and it is not surprising that we could only confirm one of them in the follow-up analysis.

TABLE 3 Haplotypes with suggested associations with a colorectal cancer syndrome

Chromosome locus	Analysis	Haplotype	F_A	F_U	OR	*P*	SNP		Genome position		Haplotype "group"[b]
							Start	Stop	Start	Stop	
1q31.3-32.1	WIN5084	AGGGAGAAGAAAA	0.036	0.017	2.2	.000018	rs7538519	rs10921881	195386897	195606741	H1
	WIN4056	GACGCAGGCG	0.032	0.015	2.2	.00_024	rs3861923	rs9663039	199116609	199242683	H2
	WIN4669	AGGGAGAAGAAA	0.036	0.018	2.0	.000027	rs7538519	rs10494727	195386897	195569614	H1
	WIN5911	AGGGAGAAGAAAAAA	0.038	0.019	2.1	.000034	rs7538519	rs2183858	195386897	195632577	H1
	WIN6734	AGGGAGAAGAAAAAAAG	0.038	0.019	2.1	.000034	rs7538519	rs7526474	195386897	195658978	H1
1q24.2–25.3	WIN7182	GAGGGAAA	0.021	0.0079	2.6	.0000093	rs3917651	rs12938	169600062	169660781	H3
	WIN9183	GAGGGAAAAG	0.021	0.0080	2.7	.000010	rs3917651	rs4987310	169600062	169673838	H3
	WIN8183	GAGGGAAAA	0.020	0.0077	2.7	.000010	rs3917651	rs964555	169600062	169671088	H3
	WIN8184	AGGGAAAAG	0.021	0.0079	2.6	.000016	rs1800805	rs4987310	169601281	169673838	H3
	WIN6181	AGGGAAA	0.021	0.0082	2.6	.000018	rs1800805	rs12938	169601281	169660781	H3
6q25.3–26	WIN12087	AAAGGGGGAAC	0.031	0.015	2.1	.000025	rs4252108	rs11060	161137523	161173946	H4
	WIN19681	GGGCAAGGGCAAAAAGCA[a]	0.048	0.027	1.8	.000036	rs4708838	rs3465	160045332	160198395	H5
	WIN20790	GGGCAAGGGCAAAAAGCAG	0.051	0.030	1.8	.000042	rs4708838	rs4709368	160045332	160204127	H5
	WIN21898	GGGCAAGGGCAAAAAGCAGG	0.051	0.030	1.8	.000042	rs4708838	rs3818299	160045332	160206631	H5
	WIN23005	GGGCAAGGGCAAAAAGCAGGG	0.051	0.030	1.8	.000043	rs4708838	rs4709369	160045332	160207570	H5
18p11.2-q11.2	WIN4411	AAAAAACG	0.043	0.026	1.7	.00034	rs1010811	rs8084363	19677786	19789487	H6
	WIN3845	AAAAACG	0.043	0.026	1.7	.00035	rs4800108	rs8084363	19694044	19789487	H6
	WIN5420	AAAAGAGGGA	0.035	0.019	1.9	.00053	rs12458598	rs9953274	13238715	13317297	H7
	WIN4855	AAAAGAGGG	0.034	0.019	1.8	.00058	rs12458598	rs16940343	13238715	13309150	H7
	WIN6665	GAAGAAAAAACG	0.037	0.021	1.8	.00065	rs2046058	rs8084363	19637794	19789487	H6
Xp11.4–11.23	WIN328	AAGGAGAAAAAAAACA	0.032	0.014	2.3	.00096	rs11091252	rs10481837	44141392	46786796	H8
	WIN372	AGAGAAAAAACAGAGAGAAAA	0.54	0.45	1.4	.0014	rs17314146	rs12007876	39784805	45803368	H9
	WIN330	AGAGAAAAAACAGAGAG	0.49	0.41	1.4	.0034	rs17314146	rs7058967	39784805	44707590	H9
	WIN353	AGAGAAAAAACAGAGAGAA	0.49	0.42	1.4	.0038	rs17314146	rs5952304	39784805	45134054	H9
	WIN342	AGAGAAAAAACAGAGAGA	0.49	0.41	1.3	.0041	rs17314146	rs7054198	39784805	44794292	H9

Note: Haplotypes from the association haplotype study showing the lowest *P* value and odds ratio above 1 in the linked loci. All positions are annotated according to GRCh37.

Abbreviations: F_A, frequency affected; F_U, frequency unaffected; SNP, single nucleotide polymorphism.

[a]Family 26 carries the haplotype.

[b]Variations of the same haplotype are sorted into the same group.

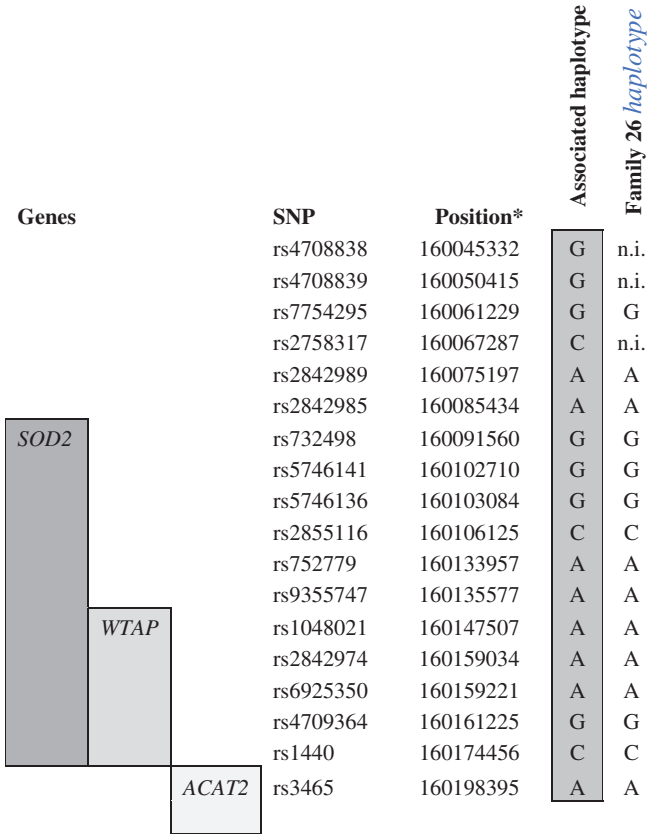

FIGURE 2 Matching haplotype on chromosome 6. The haplotype found in the association haplotype study and the corresponding haplotype in family 26, which in the linkage analysis contributed the most to the increased HLOD score in that locus. This figure also illustrates the location of known genes in this location; *SOD2* (Chr6: 160090089-160183561), *WTAP* (Chr6: 160146617-160177351), and *ACAT2* (Chr6: 160181360-16020014). All positions are annotated according to GRCh37. * Genome position on chromosome 6. Abbreviations: SNP, single nucleotide polymprphism; n.i., not informative [Color figure can be viewed at wileyonlinelibrary.com]

Family 26 had both a LOD score >1 at the locus in chromosome bands 6q25-26 and a haplotype that matched the haplotype found in the association study (Figure 2). In the exome analysis, gene-coding pathogenic variants had already been excluded.

The region of the haplotype spans the border between two topologically associating domains (TADs), and variants therein can therefore potentially disrupt gene regulation on both sides (Human, Feb 2009, hg19).[26,27] Within the haplotype region, three protein-coding genes are situated. The *SOD2* gene codes for a cellular antioxidant enzyme and has been shown to be upregulated in association with increased cancer cell migration and transition.[28] The *WTAP* gene has been suggested to be associated with the *WT1* gene, but it has no proven connection to cancer development.[29] The *ACAT2* gene product has been shown to be elevated in aggressive prostate cancer.[30] In the two TADs that border the haplotype region, 15 genes are situated. None of these genes are known to cause familial cancer syndromes as germline mutations but a majority of the genes have aberrant expression profiles in malignant cells, indicating a connection to cancer development. *FNDC1* is overexpressed in PrC; the higher the expression the more the aggressive cancer.[31] Decreased *MAS1* expression has been shown for instance in breast cancer cells and it is considered a proto-oncogene.[32] Somatic mutations in the *IGF2R* gene have been found in hepatocellular carcinoma, and *AIRN* is an imprinted gene that overlaps the *IGF2R* gene.[29] The expression of solute carrier transporters, such as *SLC22A1*, *SLC22A2*, and *SLC22A3*, has a connection to the characteristics of pancreatic cancer.[33] Aberrant LPA signaling has been described in cancer initiation and metastasis.[34] Plasmin, coded by the *PLG* gene, is involved in tumor proliferation, migration, and metastasis in several cancer types.[35]

In conclusion, in this study we suggest five new CRC, PrC, and GC syndrome susceptibility loci in chromosome bands 1q31-32, 1q24-25, 6q25-26, 18p11-q11, and Xp11. In the follow-up association haplotype study of the linked loci, we could verify a haplotype in chromosome band 6q25 in one family, who also contributed the most to the increased LOD score in that region. We therefore consider it possible that the 6q25 locus is associated with a syndrome including CRC and PrC. No proven pathogenic mutation in a colorectal, gastric, or prostate malignancy gene was detected using exome analysis within the families, contributing the most to the increased HLOD score in these regions.

ACKNOWLEDGMENTS

We thank all patients for their contribution. Genotyping was performed by the SNP&SEQ Technology Platform in Uppsala, supported by Uppsala Universitet, Akademiska Sjukhuset, Science for Life Laboratory, and the Swedish Research Council (Contracts 80576801 and 70374401). We acknowledge the possibilities to perform this study provided by the Swedish Twin Registry.

CONFLICT OF INTERESTS

The authors claim no conflict of interest. Colorectal Transdisciplinary Study (CORECT): The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the CORECT Consortium, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CORECT Consortium.

ORCID

Kristina Lagerstedt-Robinson https://orcid.org/0000-0001-9848-0468

REFERENCES

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cance —analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78-85. https://doi.org/10.1056/NEJM200007133430201.
2. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med*. 2003;348(10):919-932. https://doi.org/10.1056/NEJMra012242.
3. Al-Tassan NA, Whiffin N, Hosking FJ, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for

colorectal cancer. *Sci Rep*. 2015;5:10442. https://doi.org/10.1038/srep10442.
4. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5):759-767.
5. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell*. 1996;87(2):159-170. https://doi.org/10.1016/S0092-8674(00)81333-1.
6. Peltomäki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet*. 2001;10(7):735-740.
7. Rex DK, Ahnen DJ, Baron JA, et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol*. 2012;107(9):1315-1329. https://doi.org/10.1038/ajg.2012.161.
8. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*. 2007;50(1):113-130. https://doi.org/10.1111/j.1365-2559.2006.02549.x.
9. Forsberg AM, Hagel E, Jaramillo E, Rubio CA, Björck E, Lindblom A. Predicting outcome in colonoscopic high-risk surveillance. *Anticancer Res*. 2015;35(9):4813-4819.
10. Risio M. The natural history of adenomas. *Best Pract Res Clin Gastroenterol*. 2010;24(3):271-280. https://doi.org/10.1016/j.bpg.2010.04.005.
11. Hiripi E, Lorenzo Bermejo J, Sundquist J, Hemminki K. Association of colorectal adenoma with other malignancies in Swedish families. *Br J Cancer*. 2008;98(5):997-1000. https://doi.org/10.1038/sj.bjc.6604276.
12. Suzuki A, Koide N, Takeuchi D, et al. Prevalence of synchronous colorectal neoplasms in surgically treated gastric cancer patients and significance of screening colonoscopy. *Dig Endosc*. 2014;26(3):396-402. https://doi.org/10.1111/den.12156.
13. Lee SS, Jung WT, Kim CY, et al. The synchronous prevalence of colorectal neoplasms in patients with stomach cancer. *J Korean Soc Coloproctol*. 2011;27(5):246-251. https://doi.org/10.3393/jksc.2011.27.5.246.
14. Forsberg A, Keränen A, VON Holst S, et al. Defining new colorectal cancer syndromes in a population-based cohort of the disease. *Anticancer Res*. 2017;37(4):1831-1835. https://doi.org/10.21873/anticanres.11518.
15. Robinson KL, Liu T, Vandrovcova J, et al. Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *J Natl Cancer Inst*. 2007;99(4):291-299. https://doi.org/10.1093/jnci/djk051.
16. Kontham V, von Holst S, Lindblom A. Linkage analysis in familial non-Lynch syndrome colorectal cancer families from Sweden. *PLoS One*. 2013;8(12):e83936. https://doi.org/10.1371/journal.pone.0083936.
17. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*. 2010;26(5):589-595. https://doi.org/10.1093/bioinformatics/btp698.
18. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/. Accessed March 14, 2019.
19. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498. https://doi.org/10.1038/ng.806.
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603.
21. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. https://doi.org/10.1038/nature19057.
22. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. https://doi.org/10.1038/nature15393.
23. Ensembl genome browser 87. http://www.ensembl.org/index.html. Accessed January 6, 2017.
24. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. https://doi.org/10.1086/519795.
25. Devine E. The SNP&SEQ Technology Platform - Uppsala University, Sweden http://snpseq.medsci.uu.se/. Accessed March 14, 2019.
26. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996-1006. https://doi.org/10.1101/gr.229102.
27. UCSC Genome Browser Home. https://genome.ucsc.edu/. Accessed March 14, 2019.
28. Yi L, Shen H, Zhao M, et al. Inflammation-mediated SOD-2 upregulation contributes to epithelial-mesenchymal transition and migration of tumor cells in aflatoxin G1-induced lung adenocarcinoma. *Sci Rep*. 2017;7(1):7953. https://doi.org/10.1038/s41598-017-08537-2.
29. OMIM - Online Mendelian Inheritance in Man. https://www.omim.org/. Accessed March 14, 2019.
30. Saraon P, Trudel D, Kron K, et al. Evaluation and prognostic significance of ACAT1 as a marker of prostate cancer progression. *Prostate*. 2014;74(4):372-380. https://doi.org/10.1002/pros.22758.
31. Das DK, Naidoo M, Ilboudo A, et al. miR-1207-3p regulates the androgen receptor in prostate cancer via FNDC1/fibronectin. *Exp Cell Res*. 2016;348(2):190-200. https://doi.org/10.1016/j.yexcr.2016.09.021.
32. Luo Y, Wong C-J, Kaz AM, et al. Differences in DNA methylation signatures reveal multiple pathways of progression from adenoma to colorectal cancer. *Gastroenterology*. 2014;147(2):418-429.e8. https://doi.org/10.1053/j.gastro.2014.04.039.
33. Mohelnikova-Duchonova B, Brynychova V, Hlavac V, et al. The association between the expression of solute carrier transporters and the prognosis of pancreatic cancer. *Cancer Chemother Pharmacol*. 2013;72(3):669-682. https://doi.org/10.1007/s00280-013-2246-2.
34. Willier S, Butt E, Grunewald TGP. Lysophosphatidic acid (LPA) signalling in cell migration and cancer invasion: a focussed review and analysis of LPA receptor gene expression on the basis of more than 1700 cancer microarrays. *Biol Cell*. 2013;105(8):317-333. https://doi.org/10.1111/boc.201300011.
35. Zhao S, Dorn J, Napieralski R, et al. Plasmin(ogen) serves as a favorable biomarker for prediction of survival in advanced high-grade serous ovarian cancer. *Biol Chem*. 2017;398(7):765-773. https://doi.org/10.1515/hsz-2016-0282.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Wallander K, Liu W, von Holst S, et al. Genetic analyses supporting colorectal, gastric, and prostate cancer syndromes. *Genes Chromosomes Cancer*. 2019;58:775–782. https://doi.org/10.1002/gcc.22786