Evidence for genetic association between chromosome 1q loci and predisposition to colorectal neoplasia

This manuscript was initially published online in the wrong issue of British Journal of Cancer. This has now been rectified and a corrigendum has been published.

Stephanie A Schubert1, Dina Ruano1, Fadwa A Elsayed1, Arnoud Boot1, Arantza Farina Sarasqueta1, Bruce Wolffenbuttel2, Melanie M van der Klauw2, Jan Oosting1, Carli M Tops3, Ronald van Eijk1, Hans FA Vasen4, Rolf HAM Vossen5, Maartje Nielsen3, Sergi Castellvi-Bel6, Clara Ruiz-Ponte7, Ian Tomlinson8, Malcolm G Dunlop9, Pavel Vodicka10, Juul T Wijnen3, Frederik J Hes3, Hans Morreau1, Noel FCC de Miranda1, Rolf H Sijmons11 and Tom van Wezel*,1

1Department of Pathology, Leiden University Medical Center, Leiden University, Leiden 2300 RC, The Netherlands; 2Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands; 3Department of Clinical Genetics, Leiden University Medical Center, Leiden University, Leiden 2300 RC, The Netherlands; 4Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden University, Leiden 2300 RC, The Netherlands; 5Department of Human Genetics, Leiden University Medical Center, Leiden University, Leiden 2300 RC, The Netherlands; 6Department of Gastroenterology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Catalonia 08036, Spain; 7Fundación Pública Galega de Medicina Xéneica (FPGMX)-SERGAS, Grupo de Medicina Xéneica-USC, Instituto de Investigación Sanitaria de Santiago (IDIS), Centro de Investigación en Red de Enfermedades Raras (CIBERER), Santiago de Compostela 15706, Spain; 8Oxford Centre for Cancer Gene Research, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 9Colon Cancer Genetics Group, MRC Human Genetics Unit, The University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK; 10Institute of Experimental Medicine, Institute of Biology and Medical Genetics, Prague 142 00, Czech Republic and 11Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen 9700 RB, The Netherlands

Background: A substantial fraction of familial colorectal cancer (CRC) and polyposis heritability remains unexplained. This study aimed to identify predisposing loci in patients with these disorders.

Methods: Homozygosity mapping was performed using 222,563 SNPs in 302 index patients with various colorectal neoplasms and 3367 controls. Linkage analysis, exome and whole-genome sequencing were performed in a family affected by microsatellite stable CRCs. Candidate variants were genotyped in 10,554 cases and 21,480 controls. Gene expression was assessed at the mRNA and protein level.

Results: Homozygosity mapping revealed a disease-associated region at 1q32.3 which was part of the linkage region 1q32.2–42.2 identified in the CRC family. This includes a region previously associated with risk of CRC. Sequencing identified the p.Asp1432Glu variant in the MIA3 gene (known as TANGO1 or TANGO) and 472 additional rare, shared variants within the linkage region. In both cases and controls the population frequency was 0.02% for this MIA3 variant. The MIA3 mutant allele showed predominant mRNA expression in normal, cancer and precancerous tissues. Furthermore, immunohistochemistry revealed increased expression of MIA3 in adenomatous tissues.

Conclusions: Taken together, our two independent strategies associate genetic variations in chromosome 1q loci and predisposition to familial CRC and polyps, which warrants further investigation.

*Correspondence: Dr T van Wezel; E-mail: t.van_wezel@lumc.nl

Received 21 March 2017; revised 31 May 2017; accepted 30 June 2017; published online 25 July 2017
Family history is a major risk factor in colorectal cancer (CRC) aetiology. Familial aggregation of disease is observed in up to 25% of the cases and estimates from twin studies indicate that inherited genetic factors contribute to the development of CRCs in approximately 15% of cases (Kerber et al., 2005; Mucci et al., 2016). It is generally accepted that a substantial proportion of CRC incidence is due to predisposition to non-malignant CRC lesions (e.g., adenomas, serrated lesions). Indeed, first-degree relatives of patients with colorectal adenomas have an increased risk of developing colorectal adenocarcinomas (Johns and Houlston, 2001; Tuohy et al., 2014).

Only up to 14% of CRC predisposition can be attributed to the known Mendelian-inherited syndromes and common gene variants, suggesting that a substantial fraction of this heritability remains unexplained (Aaltonen et al., 2007; Dunlop et al., 2013; Jiao et al., 2014). In support of this, linkage studies in CRC families have identified disease-associated chromosomal areas where predisposing genes have not yet been identified (Wiesner et al., 2003; Kemp et al., 2006; Neklason et al., 2008; Middeldorp et al., 2010). Similarly, some regions, such as chromosome 1q41 (rs6691170, rs6687758 and rs11118883), identified in genome-wide association studies (GWAS) still have not led to the identification of functional CRC variants (Houlston et al., 2010; Spain et al., 2012; Zhang et al., 2014). Furthermore, one study measuring runs of homozygosity in search for recessive disease-causing loci, did not identify novel CRC predisposing loci (Spain et al., 2009). In recent years, advances in sequencing technologies have made it possible to identify germline susceptibility loci that were previously not encountered in linkage and association studies. For example, using a combination of sequencing and linkage studies germline mutations were identified in two DNA-polymerase genes (POLE and POLD1) (Palles et al., 2013).

In this study, we searched for germline variants involved in colorectal neoplasm susceptibility using two independent genomic strategies. Homozygosity mapping analysis was conducted in a cohort of 302 index patients with various colorectal neoplasms, ranging from few to hundreds of polyps and 3367 controls. In parallel, a combination of linkage analysis and sequencing was performed in individuals from a large family with an extensive history of microsatellite stable CRC. Both strategies identified an association of chromosome 1q loci with these diseases.

**MATERIALS AND METHODS**

**Ethics.** The study was approved by the Medical Ethical Committee of the Leiden University Medical Center, Leiden, The Netherlands (protocol P01-019). Informed consent was given for further research at initial blood withdrawal for the homozygosity mapping cohort, family 68 and the MIA3 genotyping cohorts. Anonymised tissue samples were handled according to the medical ethical guidelines described in the Code Proper Secondary Use of Human Tissue established by the Dutch Federation of Medical Sciences (https://www.federa.org/) and informed consent was waived for the use of these samples by the Medical Ethical Committee (protocol P01-019).

**CytoSNP array genotyping.** A cohort of patients with multiple adenomas, previously described by (Hes et al., 2014), and patients with <10 adenomas or a phenotype with predominantly serrated lesions, were included in this study. The clinical characteristics of the cohort are presented in Supplementary Table 1. All cases were negative for known pathogenic variants in the MMR genes, APC (including mosaic APC), POLE, POLD1, MUTYH or NTHLI. A total of 343 cases were genotyped using the Illumina HumanCytoSNP-12 v2.0 or v2.1 BeadChips (Illumina Inc., San Diego, CA, USA). For controls we used genotypes of 3367 individuals from the LifeLines cohort study (Stolk et al., 2008; Dolmans et al., 2011). A total of 222 563 SNPs in 302 cases and 3367 controls passed quality control. The quality control procedure is described in Supplementary Methods.

**Homozygosity mapping.** SNP selection and homozygosity mapping analysis is described in Supplementary Methods. In total 78 582 SNPs were included in the analysis. Runs of homozygosity were compared between cases and controls using the Fisher’s exact test ($P<1.2 \times 10^{-8}$).

**Genome-wide association study.** A quantile–quantile (Q–Q) plot of the observed and expected $P$-values indicated differential genotyping between cases and controls. A genomic inflation factor (based on median $\chi^2$) $\lambda=1.23378$ was applied to correct for the population bias between cases and controls. Genotype prevalence of 222 563 SNPs was compared between cases and controls using a $\chi^2$ allelic test with 1 degree of freedom. Odds ratio and its 95% confidence interval was also computed. Multiple test correction was done using false discovery rate estimation (Benjamini and Yekutieli, 2001).

**Family 68 characterisation.** Clinicopathological data of family 68 family members were obtained from the national registry of The Netherlands Foundation for the Detection of Hereditary Tumours (http://www.stoet.nl) and are shown in Supplementary Table 2. Peripheral blood was collected from 16 members of family 68. DNA was extracted using standard techniques. Index patient 68-01 had been screened and was found to be negative for known pathogenic variants in the MMR genes, APC, POLE, POLD1, MUTYH and NTHLI.

**Linkage analysis.** Fifteen family members of family 68, including one unaffected spouse, were genotyped using the GeneChip Mapping 10 K Xba 142 array (Affymetrix Inc., Santa Clara, CA, USA). Data was previously described (Middeldorp et al., 2010). SNPs with a call rate below 90% were excluded from further analysis. A total of 9131 SNPs had a valid annotation. Genotypes incompatible with the family relations were removed using pedcheck. Non-parametric linkage analysis was performed with MERLIN (Abecasis et al., 2002) using S-all scoring (Whittemore and Halpern, 1994) with the exponential model (Kong and Cox, 1997). Patients diagnosed with CRC before 60 years or with colorectal polyps before 55 years were considered affected. Other family members were classified as unaffected.

**Exome and whole-genome sequencing.** Exome sequencing was performed on family 68 DNA samples (68-01, 68-08, 68-11, 68-15 and 68-17) using HiSeq (Illumina) or Complete Genomics (Complete Genomics Inc., Mountain View, CA, USA) sequencing technology (Carnevali et al., 2012) as described in Supplementary Methods. Samples (68-07, 68-11 and 68-15) were sent to Macrogen (Macrogen Europe, Amsterdam, The Netherlands) for whole-genome sequencing (Supplementary Methods). Variants with a population frequency $>0.01$ (in 1000Genomes, ExAC, ESP or GoNL databases) were excluded. For the analysis of the exome sequencing data, we selected variants predicted to be coding or affecting splice-sites. In the whole-genome analysis, we prioritised the non-coding variants which were present within the linkage peak based on regions under strong negative selection, as previously described (Khurana et al., 2013).

**MIA3 p.Arg1432Glu genotyping.** The MIA3 c.4296T>A, p.Asp1432Glu variant was validated by standard bidirectional Sanger sequencing. A KASPar assay was designed for the MIA3 p.Asp1432Glu variant by KBioscience (LGC Genomics, Hoddesdon, UK). Leucocyte or normal colon tissue derived DNA was screened of 1477 CRC patients, 445 polyposis patients and 1604 population-based controls (anonymous blood donors) from The Netherlands and 1098 CRC patients and 1459 healthy controls.
from the Czech Republic (Pardini et al., 2013). All carriers were validated by Sanger sequencing and the existence of a common ancestor was excluded using STR profiling (Supplementary Methods). Additionally, MIA3 genotypes were extracted from a European cohort of 7534 CRC patients and 18417 healthy controls that were genotyped using OmniExpressExome BeadChip 8v1.1 or 8v1.20 (Illumina) (Timofeeva et al., 2015).

**Somatic MIA3 mutation analysis.** Tumour DNA of 47 microsatellite stable CRCs was screened for somatic MIA3 mutations (Supplementary Table 3). Paired tumour-normal DNA samples of MIA3 p.Asp1432Glu carriers (68-1 and both carriers identified with screening) were screened for second hits in the MIA3 gene. Sequencing was performed with a custom multiplex PCR on the Ion Torrent Personal Genome Machine sequencer (Life Technologies, Grand Island, NY, USA) as described in Supplementary Methods. Synonymous and intronic variants were excluded as well as variants present in population frequency databases (in 1000Genomes, ExAC, ESP or GoNL).

For comparison, the somatic mutation data from 915 colorectal adenocarcinoma exomes (Muzny et al., 2012; Seshagiri et al., 2012; Giannakis et al., 2016) was retrieved from The cancer genome atlas (TCGA) using cBioPortal (Cerami et al., 2012; Gao et al., 2013).

**Quantitative real-time reverse transcriptase PCR.** mRNA was isolated from fresh frozen tissue samples of 18 normal tissues, 32 colorectal adenomas and 18 colorectal adenocarcinomas (Supplementary Table 4). mRNA isolation and quantitative real-time reverse transcriptase PCR were performed as previously described (Lips et al., 2008; van Roon et al., 2013). MIA3 mRNA expression of the full-length isoform (primer sequences available upon request) was normalised to housekeeping genes CPSF6 and HNRNPM (van Roon et al., 2013). Group differences were tested with one-way ANOVA.

**Allele-specific expression.** Formalin-fixed, paraffin-embedded tissue blocks of MIA3 p.Asp1432Glu carriers were collected (caecum carcinoma and normal tissue; two serrated lesions, two adenomas and normal colon tissue). Total nucleic acid was extracted using the Tissue Preparation System (Siemens Healthcare Diagnostics, Tarrytown, NY, USA) (van Eijk et al., 2013). Nucleic acid was treated with DNasel (Fermentas, Thermo Fisher Scientific, Waltham, MA, USA) and converted into cDNA using the High-Capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA, USA). Allele-specific expression analysis of the MIA3 variant was performed using 2 μl 25× diluted cDNA as template for the KASPar assay. Paired DNA was analysed to determine baseline amplification differences between the two alleles. Using the Cq values obtained for both alleles, the allelic dosage was calculated similarly to the Pfaffl method for relative gene expression (Pfaffl, 2001).

**Immunohistochemistry.** Formalin-fixed, paraffin-embedded tissue blocks were available of 82 patients from the FACTS study (Hennink et al., 2015) (which included 18 serrated lesions and 97 adenomas with low-grade dysplasia) and a series of 15 anonymous colorectal adenocarcinomas. Formalin-fixed, paraffin-embedded and fresh frozen tissue sections of MIA3 p.Asp1432Glu carriers were stained. MIA3 immunohistochemistry was performed as described in Supplementary Methods. The intensity of MIA3 cytoplasmic staining was classified as negative-weak or moderate-strong. Normal mucosa was analysed when present in the tissue sections. MIA3 gradient expression was scored: maximal staining at the base of the crypts and decreasing towards the lumen (top-down), maximal staining at the lumen and decreasing towards the base of the crypts (down-up) or homogeneous staining across lumen and crypt (no gradient). Differences between tissue types were compared using a Fisher’s exact test.

**RESULTS**

Genotyping and homozygosity mapping in colorectal neoplasm cases. We genotyped germline DNA samples from index patients affected with colorectal neoplasms ranging from few to hundreds of polyps (polyposis), with and without CRC. After sample and SNP quality control procedures and removal of ethnic outliers, the

---

**Figure 1.** Overrepresentation of runs of homozygosity located on chromosome 1q32.3 in patients with colorectal neoplasms. (A) Frequency of the runs of homozygosity on chromosome 1; cases (blue) and controls (pink). Overrepresented homozygosity in cases is indicated with a red box. (B) Overlapping runs of homozygosity of six cases (blue bars) at chromosomal locus 1q32.3.
Clinical characterisation and linkage analysis in family 68. In parallel to the homozygosity mapping analysis we performed linkage analysis in a family not included in the homozygosity mapping cohort. Updated clinical follow-up information warranted the reanalysis of a previously described Amsterdam criteria fulfilling family with microsatellite stable CRCs without a known causative mutation (Middeldorp et al., 2010). Fifteen individuals of this family (68-01 until 68-15) were genotyped using Affymetrix 10 K SNP arrays (Figure 2). For non-parametric linkage analysis only individuals presenting with early-onset colorectal neoplasms (>1 adenoma before 55 years and/or CRC before 60 years) were considered affected. All affected family members developed less than 10 polyps. The clinical presentation of the late-onset patients included the diagnosis of CRC in an elderly patient at 88 years of age (68-13) and the identification of 1–2 adenomas during periodic colonoscopy screenings in three relatives, aged 56–60 years (68-02, 68-06 and 68-09). Furthermore, the linkage analysis included one family member without a history of polyps or cancer at 59 years of age and one unaffected spouse. We identified a region of suggested linkage (logarithm of odds (LOD) > 2.5; maximum LOD score = 2.75) on chromosome 1q32.2-q42.2 (Figure 3). No additional linkage regions were identified (LOD > 0.8). The linkage region spans 21.1 Mb (chr1:210 940 342–232 031 815) and overlaps with the region of homozygosity described above, located on chromosome 1q32.3. The whole region is predicted to contain 137 coding genes and 48 small non-coding RNAs (UCSC gene database). No individual SNPs, including chromosome 1q SNPs rs11119874 and rs1338348, was identified in six patients diagnosed with multiple polyps before 56 years of age; five out of six had >10 neoplastic lesions (Supplementary Table 5). This region contains 10 coding genes and 2 small non-coding RNAs (UCSC gene database). No individual SNPs, including chromosome 1q SNPs rs11118883 and rs6687758, reached global significance in a GWAS analysis utilising the same SNPs.

Germline sequencing in family 68. To further investigate the chr1q loci identified with homozygosity mapping and linkage analysis, we focussed on the largest region, the linkage region in family 68. In search of disease-associated variants, exome sequencing was performed on DNA from five affected family members (68-1, 68-8, 68-11, 68-15 and 68-17). Exonic or predicted splice-site variants with a maximum population frequency below 0.01 were selected. Three variants were shared amongst these five individuals (Table 1), of which only one was located within the linkage peak: the c.4296T>A, p.Asp1432Glu variant in Melanoma Inhibitory Activity protein 3 gene (MIA3, also known as TANGO1 or TANGO). Additional variants identified within the linkage peak were not shared between the individuals (Supplementary Table 6). Segregation analysis, using Sanger sequencing, showed that all family members presenting with early-onset colorectal neoplasms were carriers of the MIA3 p.Asp1432Glu variant, whereas one unaffected relative (68-04), one patient with a single polyp at 60 years of age (68-9) and one unaffected spouse (68-14) were non-carriers. The MIA3 variant was present in three late-onset patients (68-02, 68-06, 68-13).

To investigate the presence of variants in non-coding regions on chromosome 1q32.2-q42.2, whole-genome sequencing was performed in three family members (68-7, 68-11 and 68-15). Of the known CRC risk alleles located on chromosome 1q (rs6691170, rs6687758 or rs11118883) all individuals carried the non-risk alleles with one exception, individual 68-11 was heterozygous for the rs6691170 allele. Next, heterozygous variants shared amongst all three individuals and with a maximum population frequency of 0.01, were extracted from within the linkage peak region. A total of
473 variants were identified. 51% were single nucleotide variants and 49% insertion or deletions (81% in genetic repeat sequences).

In total 33 variants were located within the homozygosity mapping region located on chromosome 1q32.3. Within 1 Mb of the MIA3 gene 45 variants were identified (chr1:221 791 444–223 841 354). In addition, one exonic variant was found (OBSCN: NM_001098623 c.4523T>A, p.Val1508Asp), which was not covered in the exome sequencing. Subsequently, we investigated variants located in regions under strong negative selection, i.e. non-coding regions harbouring an increased fraction of rare variants comparable to the coding region, which are subdivided into ultrasensitive (e.g., BRF2-binding site) and sensitive regions (binding sites of some chromatin and general transcription factors). One variant, rs529680452, was identified in a sensitive non-coding element; in the binding domain of histone acetyltransferase KAT2A located in the 5' untranslated region of the DTL gene (Supplementary Table 7). Thus, the MIA3 p.Asp1432Glu variant seemed the most promising candidate for the colorectal neoplasm predisposition in this family and was selected for further investigation.

Prevalence of the MIA3 p.Asp1432Glu variant. The MIA3 p.Asp1432Glu variant is present at low frequencies in public genotype databases; 0.01% the in ESP database (1/5997), 0.05% in the ExAC database (55/59 662) and 0.1% in the Genome of The Netherlands project (1/498). We performed or collected additional genotyping data of 10 554 cases and 21 480 controls. Analysis over all cohorts revealed an allele frequency of 0.02% in both cases and controls; separate frequencies per cohort are shown in Supplementary Figure 1. KASPar genotyping of the Dutch CRC cohort identified two additional carriers of the p.Asp1432Glu variant. No carriers were found in the cases used for homozygosity mapping.

Clinical characterisation MIA3 p.Asp1432Glu carriers. STR profiling provided no evidence of a recent common ancestor between the 68-01 and the other two carriers. One patient has a positive family history of both CRC and adenomas and has a confirmed germline deletion in MSH2. This patient developed a microsatellite instability-high (MSI-H) caecum tumour at age 44.

Table 1. Variants present in all five affected family members identified with exome sequencing

| Gene   | Chromosome band | GenBank accession number | mRNA change | Protein change | Reference SNP cluster ID |
|--------|-----------------|--------------------------|-------------|----------------|-------------------------|
| MIA3   | 1q41            | NM_198551                | c.4296 T>A  | p.Asp1432Glu   | rs199885504             |
| SIPA1L2| 1q42.2          | NM_020808                | c.4966 G>A  | p.Gly166Arg    | —                       |
| FAM134A| 2q35            | NM_024293                | c.1156_1161del | p.386_387del   | —                       |

Figure 4. mRNA and protein expression of MIA3 in colon tissues. (A) mRNA expression of MIA3 (relative to housekeeping genes CPSF6 and HNRNPM) in colonic tissues. (B) The p.Asp1432Glu variant was higher expression than the wild-type allele in both normal colonic tissue and colon lesions. (C) A significant increase in protein expression is observed in adenomas and a significant decrease in carcinomas, compared to normal colonic tissue sections. (D) Gradient expression of MIA3 protein was more frequent in serrated lesions and adenomas compared to carcinomas. Serrated lesions show higher expression basally, while adenomas had higher expression on the luminal side of the polyp. *P<0.05, **P<0.01.
The other patient had no germline mutations in CRC genes and no known family history of disease. The patient developed microsatellite stable rectal cancer at 40 years and underwent a hemicolectomy of the left-sided colon and rectum at age 56 after developing >20 polyps (both adenomas and serrated lesions).

**MIA3 somatic mutations.** We performed targeted somatic mutation screening of MIA3 in tumour DNA from p.Asp1432Glu MIA3 carriers. In the MSI-H tumour a somatic mutation was present (c.5570G>A, p.Gly1857Asp). No second hits in MIA3 were detected in the other tumour samples. In addition, MIA3 was screened in 47 familial microsatellite stable CRCs. Somatic non-synonymous MIA3 mutations were observed in 6.4% of the tumours (Supplementary Table 8). Our findings are in line with the somatic mutation frequency of 4.5% in 915 colorectal adenocarcinomas from TCGA.

**MIA3 mRNA and allelle-specific expression.** Using quantitative real-time PCR, MIA3 expression was quantified in mRNA isolated from normal tissues, colorectal adenomas and colorectal adenocarcinomas. High variability in expression was observed within each group and showed no significant difference in expression between the groups (P = 0.684; Figure 4A). We investigated the allele-specific expression of the MIA3 variant in tissues derived from p.Asp1432Glu MIA3 carriers. mRNA expression of the mutant allele was three times higher in normal tissue samples and more than four times higher in colorectal neoplasms (Figure 4B).

**MIA3 protein expression.** To better understand the potential role of MIA3, we examined MIA3 expression in precancerous colorectal lesions and colorectal carcinoma tissue sections. A general decrease in MIA3 expression was observed in adenocarcinomas compared to normal mucosa (13 vs 46% moderate-strong staining; P = 0.04; Figure 4C). Notably, higher MIA3 expression was seen in colorectal polyps, 67% of serrated lesions and 70% of adenomas showed moderate-strong staining (P = 0.18, P = 0.006, respectively, compared to normal mucosa). A gradient in MIA3 expression was observed in polyps (representative tissue sections in Supplementary Figure 2), while no gradient was present in normal mucosa. In serrated lesions MIA3 expression was highest at the base of the crypts and decreased towards the lumen (down-top gradient), while in adenomas the inverse gradient was observed (top-down gradient) (Figure 4D).

In tissue sections from p.Asp1432Glu carriers, the MIA3 expression was highly variable but not distinctive from non-carriers. These results mainly show high inter- and intra-patient heterogeneity, independent of the MIA3 p.Asp1432Glu germline variant.

**DISCUSSION**

To date, genetic association studies have reported many SNPs associated with either CRC or multiple adenoma susceptibility, including rs6691170, rs6687758, rs11118883 located on chromosome 1q41 (Spain et al, 2012; Theodoratou et al, 2012; Montazeri et al, 2016). In contrast, homozygosity mapping studies have been less successful in the identification of predisposition loci in CRC and other types of cancers, indicating that recessive syndromes might be less frequent in genetic predisposition to cancer (Spain et al, 2009; Enciso-Mora et al, 2010; Hosking et al, 2011; Sud et al, 2015). In this study we found a recurrent region of homozygosity located on chromosome 1q32.3 (chr1:211 265 284–212 357 017) in patients with a mixed early-onset colorectal neoplasm phenotype. For this study we included patients with a large spectrum of cancerous and precancerous colorectal lesions, as known hereditary CRC syndromes also have a highly variable clinical presentation. None of the runs of homozygosity or single SNPs in our GWAS analysis reached significance with the genome-wide thresholds, which mainly reflects a lack of power and underscores the need for larger sample sizes. In a highly overlapping cohort, two variants significantly associated with disease, rs3802842 (chr11q23) and rs4779584 (chr15q13), however, only 16 SNPs were tested (Hes et al, 2014). Hence, despite adjustment for multiple testing the association between early-onset colorectal neoplasm risk and chromosome 1q32.3 in susceptibility to multiple colorectal neoplasms warrants further investigation.

Furthermore, a linkage peak was identified in 15 individuals from a large family with an extensive history of microsatellite stable CRC and polyps located on chromosome 1q32.2–42.2. As an individual finding the identification of the linkage region in a single family or the region of homozygosity in the colorectal neoplasm cohort might not provide compelling evidence for the involvement of this region. However, together the linkage region, the 21.1 Mb region of homozygosity and previously described predisposing and protective CRC risk alleles (Spain et al, 2012) located in a highly overlapping region are suggestive of a broad-based association of chr1q and CRC predisposition. Therefore we decided to further investigate this association by focussing on the broadest, most inclusive region, which was the linkage region identified in family 68.

Although the family presented with a dominant pattern of inheritance, we proceeded with the screening of rare genetics variants in family 68 looking into heterozygous, homozygous and compound heterozygous variants within the linkage region. Based on exome sequencing results, the most likely candidate variant for predisposition was the heterozygous MIA3 c.4296T>A; p.Asp1432Glu variant. MIA3 (also known as TANGO or TANGO1) is a cargo receptor localised to the endoplasmic reticulum and facilitates the trafficking of collagens to the Golgi apparatus (Wilson et al, 2011; Ishikawa et al, 2016). The MIA3 p.Asp1432 amino acid is located in the cytoplasm for predisposition was the heterozygous MIA3 c.4296T>A; p.Asp1432Glu variant. MIA3 (also known as TANGO or TANGO1) is a cargo receptor localised to the endoplasmic reticulum and facilitates the trafficking of collagens to the Golgi apparatus (Wilson et al, 2011; Ishikawa et al, 2016). The p.Asp1432 amino acid is located in the cytoplasm in the first out of two coiled-coiled domains of MIA3 (amino acids 1211–1440) (Saito et al, 2011). This region may interact with still unidentified intermediates of this transport carrier complex and we hypothesised that missense variants may stabilise or destabilise these protein interactions. MIA3 was shown to promote angiogenesis and lymphangiogenesis in oral squamous cell carcinoma (Sasahira et al, 2014). Furthermore, patients with MIA3-positive squamous cell carcinomas have a significantly shorter disease-free survival than patients without MIA3 expression (Sasahira et al, 2016). However, down-regulation was observed in melanomas, hepatocellular and colon carcinomas (Arndt and Bosserhoff, 2006; 2007; Sasahira et al, 2014). To our knowledge we are the first to describe an increased protein expression in colorectal adenomas, which might indicate a tumourigenic role of MIA3 in the development of these precancerous lesions. In addition, our data showed a tissue gradient in adenomas and the inverse gradient in serrated lesions, possibly correlated to the proliferating cells in these polyps. At the mRNA level no differences were observed between adenomas, carcinomas and normal colon tissue.

No differential protein expression was observed in tissue from carriers and non-carriers of the MIA3 p.Asp1432Glu variant, although only a limited number of tissue blocks were available from carriers. In addition, genotyping data did not support the involvement of this MIA3 variant in CRC predisposition, based on similar allele frequencies in cases and controls. The observed allelic imbalance of the p.Asp1432Glu MIA3 allele in carriers could possibly be explained by the presence of regulatory variants in linkage, leading to the up-regulation of the mutant allele compared to the wild-type allele in these individuals.

To investigate the possibility of non-coding variants underlying CRC predisposition in family 68, whole-genome sequencing was performed. We focussed on heterozygous variants within the
linkage region due to the predicted dominant pattern of inheritance in this family. In total 473 heterozygous, shared, rare variants were identified. One additional coding variant, OBSCN p.Arg4213His, was identified in a region not covered by the exome sequencing analysis. The OBSCN gene codes for obscurins, giant cytoskeletal Rho-guanine nucleotide exchange factor proteins, that interact with cytoskeletal calmodulin and titin (Young et al., 2001). The OBSCN gene is frequently mutated in CRC and other cancers, however, to date, research lacks to support an association of germline variants and increased cancer risks (Sjoblom et al., 2006; Balakrishnan et al., 2007; Huhn et al., 2014).

As no regulatory elements of MIA3 have been described, we selected variant located between 1 Mb up- and down-stream of the MIA3 gene and identified 45 variants (Sotelo et al., 2010). None of these variants were located in regions under strong negative selection (Khurana et al., 2013), such as regions important for transcription factor binding, and therefore additional research is needed to investigate the effect of these 45 variants on MIA3 transcription. Of the 473 variants only one was present in a region under strong negative selection, a rare variant (rs529680452) in the binding domain of histone acetyltransferase KAT2A located in the 5’ untranslated region of the DTL gene. Denticileless E3 ubiquitin protein ligase homologue, encoded by the DTL gene, is upregulated in various cancers, including CRC, and it is reported to play a crucial role in carcinogenesis (Cheung et al., 2001; Pan et al., 2006; Ueki et al., 2008; Baranskin et al., 2012; Kobayashi et al., 2015). Variants in transcription factor binding sites have previously been associated with an increased risk of cancer, the most well studied are long-range enhancers of MYC (Ahmadiyeh et al., 2010; Sotelo et al., 2010). The DTL variant is one out of 33 non-coding variants located in the 1 Mb region identified with homozygosity mapping analysis on chromosome 1q32.3, which are also interesting candidates and warrant for further investigation.

Taken together, this study identifies, with two independent strategies, a genomic region on chromosome 1q which is associated with the predisposition to CRC and multiple polyps. Several other studies have pinpointed this region in CRC genetic association studies. Novel comprehensive approaches are required for the identification of functional variants predisposing to CRC, possibly affecting regulatory elements or non-coding RNAs.

ACKNOWLEDGEMENTS

This publication is supported by COST Action BM1206 and MSMT LD14050. This study was funded by grants from the Dutch Digestive Foundation (MLDS FP13-13), Dutch Cancer Society (KWF UL2010-4656), Stichting Sasha Swarttouw-Hijmans and Leids Universiteits Fonds/Gratama Stichting. This work was further supported by grants from Fondo de Investigación Sanitaria/FEDER (14/00173, 14/00340). Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Catalan Tumour Bank Network (Pla Director d’Oncologia, Generalitat de Catalunya), CERCA Program (Generalitat de Catalunya) and Agència de Gestió d’Ajuts Universitaris i de Recerca (Generalitat de Catalunya, 2014SGR255). We are sincerely grateful to the Biobank of Hospital Clinic-IDIBAPS for technical help. The work was carried out in part at the Esther Koplowitz Centre and FPGMX. CIBERehd and CIBERERD are funded by the Instituto de Salud Carlos III. We acknowledge core funding to the Wellcome Trust Centre for Human Genetics from the Wellcome Trust (090532/Z/09/Z) and the Oxford NIHR Biomedical Research Centre. We thank Rashmi Dibipersat, Nesli Ozcelik, Melanie Schrumpf and Brendy van den Akker for their technical support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R (2007) Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res* 13(1): 356–361.

Abeysri GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30(1): 97–101.

Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almeldo V, He IH, Brown M, Liu XS, Davis M, Caswell JL, Beckwith CA, Hills A, Macconnaill L, Coetzee GA, Regan MM, Freedman ML (2010) 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci USA* 107(21): 9742–9746.

Arndt S, Bosserhoff AK (2006) TANGO is a tumor suppressor of malignant melanoma. *Int J Cancer* 119(12): 2812–2820.

Arndt S, Bosserhoff AK (2007) Reduced expression of TANGO in colon and hepatocellular carcinomas. *Oncof Rep* 18(4): 885–891.

Balakrishnan A, Bleeker FE, Lamba S, Rodolfo M, Daniotti M, Scarpa A, van Tilborg AA, Leenstra S, Zanot C, Bardelli A (2007) Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic cancer. *Cancer Res* 67(8): 3545–3550.

Baranisink A, Birkenkamp-Demtroder K, Maghnouj A, Zollner H, Munding J, Klein-Scory S, Reinaech-Schick A, Schwarte-Waldhoff I, Schmiegel W, Hahn SA (2012) MiR-30a-5p suppresses tumor growth in colon carcinoma by targeting DTL. *Carcinogenesis* 33(4): 732–739.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29(4): 1165–1188.

Carnevali P, Baccash J, Halpern AL, Nazarenko I, Nilsen GB, Pant KP, The Elbers JC, Brownley A, Morenzoni M, Karpynch Y, Martin B, Ballinger DG, Drmanac R (2012) Computational techniques for human genome resequencing using mated gapped reads. *J Comput Biol* 19(3): 279–292.

Cerami E, Gao J, Dogrusoz U, Gross B, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5): 401–404.

Cheung WM, Chu AH, Chu PW, Ip NY (2001) Cloning and expression of a novel nuclear matrix-associated protein that is regulated during the retinoic acid-induced neuronal differentiation. *J Biol Chem* 276(20): 17083–17091.

Dolmans GH, Werker PM, Hennes HS, Furniss D, Festen EA, Franke L, Becker K, van der Vlies P, Wolffenbuttel BH, Timschoerlt, Sotolat MR, Nothnagel M, Franke A, Klopp N, Wichmann HE, Nurnberg P, Giele H, Ophoff RA, Wijmenga C (2011) Wnt signaling and Dupuytren’s disease. *N Engl J Med* 365(4): 307–317.

Dunlop MG, Tenesa A, Farrington SM, Balleauere S, Brewster DH, Koessler T, Pharoah P, Schafmayer C, Hampe J, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, von Holst S, Picelli S, Lindblom A, Jenkins MA, Hopper JL, Casey G, Duggan D, Newcomb PA, Abul B, Bessa X, Ruiz-Ponte C, Castellvi-Bel S, Niittymaki I, Tuapanen S, Karhu A, Aaltonen L, Zanke B, Hudson T, Gallinger S, Barclay E, Martin I, Gorman M, Carvafla-Carmona L, Walther A, Kerr D, Lubbie S, Broderick P, Chandler J, Pittman A, Penegar S, Campbell H, Tomlinson I, Houlston RS (2013) Cumulative risk impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut* 62(6): 871–881.

Enciso-Mora V, Hosking FJ, Houlston RS (2010) Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *Eur J Hum Genet* 18(8): 909–914.

Gao J, Aksoy BA, Dogrusoz U, Dresner G, Sander C, Sander SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Sander C, Schultz N (2013) Integrative analysis of complex cancero genes and clinical profiles using the cBioPortal. *Sci Signal* 6(269): pl1, Karpynch V, Martin B, Table 1.

Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, Bahl S, Cao Y, Amin-Mansour A, Yamauchi M, Sukawa Y, Stewart C,
Chromosome 1q loci in familial colorectal cancer

Geyer M, Ozenerben B, Palchik J, Peterson J, Sofia H, Thomson E (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487(7407): 330–337.

Neklasdon DW, Kerber RA, Nilsson DB, Anton-Culver H, Schwartz AG, Griffin CA, Lowry JT, Schildkraut JM, Evans JP, Tomlinson GE, Strong LC, Miller AR, Stopper JE, Finkelstein DM, Nadkarni KM, Kasten CH, Mineau GP, Burt RW (2008) Common familial colorectal cancer linked to chromosome 1q31: a genome-wide analysis. Cancer Res 68(21): 8993–8997.

Palles C, Cazier JB, Howarth KM, Jones AM, Broderick P, Rosa F, Barone E, Di Gaetano C, Slyskova J, Novotny J, Levy M, www.bjcancer.com |

Chromosome 1q loci in familial colorectal cancer

sequences of human breast and colorectal cancers. Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, R-spondin fusions in colon cancer. Haverty PM, Bourgon R, Wang W, Koeppen H, Gentleman R, Starr TK, Dijkgraaf GJ, Stinson J, Gnad F, Huntley MA, Degenhardt JD, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, Ha C, enhancers on 8q24 regulate c-Myc.

Stephens RM, Harris TJ, Munroe DJ, Wu X (2010) Long-range organisation protein 1 is a novel tumour progressive factor in oral squamous cell carcinoma. Bhawal UK, Bosserhoff AK, Kuniyasu H (2014) Transport and Golgi family in malignancies: MIA gene family members are novel, useful markers of esophageal, lung, and cervical squamous cell carcinoma. Oncotarget 7(21): 31137–31152.

Samihara T, Kita T, Nishiguchi Y, Kurihara M, Nakashima C, Bosserhoff AK, Kuniyasu H (2016) A comprehensive expression analysis of the MIA gene family in malignancies: MIA gene family members are novel, useful markers of esophageal, lung, and cervical squamous cell carcinoma. Mol Biol Cell 27(5): 2301–2308.

Katada T (2011) cTAGE5 mediates collagen secretion through interaction with TANGO1 at endoplasmic reticulum exit sites. Eur J Cancer 47: 3113–3115.

Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Sotelo J, Esposito D, Duhagon MA, Banfield K, Mehalko J, Liao H, Sasahira T, Kirita T, Nishiguchi Y, Kurihara M, Nakashima C, Bosserhoff AK, Saito K, Yamashiro K, Ichikawa Y, Erlmann P, Kontani K, Malhotra V, Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Pardini B, Rosa F, Barone E, Di Gaetano C, Slyskova J, Novotny J, Levy M, www.bjcancer.com | Attribution-Non-Commercial-Share Alike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/.

© The Author(s) named above 2017