




Article

GWAS Links New Variant in Long Non-Coding RNA *LINC02006* with Colorectal Cancer Susceptibility

Ewa E. Hennig ^{1,2,*}, Anna Kluska ², Magdalena Piątkowska ², Maria Kulecka ^{1,2}, Aneta Bałabas ², Natalia Zeber-Lubecka ^{1,2}, Krzysztof Goryca ^{2,†}, Filip Ambrozkiewicz ², Jakub Karczmarski ², Tomasz Olesiński ³, Łukasz Zyskowski ³ and Jerzy Ostrowski ^{1,2}

- ¹ Department of Gastroenterology, Hepatology and Clinical Oncology, Centre of Postgraduate Medical Education, 02-781 Warsaw, Poland; mkulecka@cmkp.edu.pl (M.K.); natalia.zeber-lubecka@cmkp.edu.pl (N.Z.-L.); jostrow@warman.com.pl (J.O.)
- ² Department of Genetics, Maria Skłodowska-Curie National Research Institute of Oncology, 02-781 Warsaw, Poland; anna.kluska@pib-nio.pl (A.K.); magdalena.piatkowska@pib-nio.pl (M.P.); aneta.balabas@pib-nio.pl (A.B.); kgoryca@gmail.com (K.G.); filip.ambrozkiewicz@pib-nio.pl (F.A.); jakub.karczmarski@pib-nio.pl (J.K.)
- ³ Department of Gastroenterological Oncology, Maria Skłodowska-Curie National Research Institute of Oncology, 02-781 Warsaw, Poland; tomasz.olesinski@pib-nio.pl (T.O.); lukasz.zyskowski@pib-nio.pl (Ł.Z.)
- * Correspondence: ehennig@cmkp.edu.pl
- † Current address: Genomics Core Facility, Centre of New Technologies, University of Warsaw, Poland.



Citation: Hennig, E.E.; Kluska, A.; Piątkowska, M.; Kulecka, M.; Bałabas, A.; Zeber-Lubecka, N.; Goryca, K.; Ambrozkiewicz, F.; Karczmarski, J.; Olesiński, T.; et al. GWAS Links New Variant in Long Non-Coding RNA *LINC02006* with Colorectal Cancer Susceptibility. *Biology* **2021**, *10*, 465. <https://doi.org/10.3390/biology10060465>

Academic Editor:
Salvatore Piscuoglio

Received: 23 April 2021
Accepted: 20 May 2021
Published: 25 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Identifying risk factors for cancer development can allow for appropriate stratification and surveillance of individuals at risk, increasing their chances of benefiting from early disease detection; however, most of the genetic factors contributing to the risk of colorectal cancer (CRC) remain undetermined. Here, we adopted a new approach for selecting index polymorphism for further validation in combination with a genome-wide association study of pooled DNA samples for CRC susceptibility variants in the Polish population. This study, including 2013 patients and controls, uncovered five susceptibility loci not previously reported for CRC. Four of identified variants were located within genes likely involved in tumor invasiveness and metastasis, suggesting that they could be markers of poor prognosis in CRC patients. Our results provide evidence that conducting association studies on small but homogenous populations can help us discover new common risk variants specific to the studied population.

Abstract: Despite great efforts, most of the genetic factors contributing to the risk of colorectal cancer (CRC) remain undetermined. Including small but homogenous populations in genome-wide association studies (GWAS) can help us discover new common risk variants specific to the studied population. In this study, including 465 CRC patients and 1548 controls, a pooled DNA samples-based GWAS was conducted in search of genetic variants associated with CRC in a Polish population. Combined with a new method of selecting single-nucleotide polymorphisms (SNPs) for verification in individual DNA samples, this approach allowed the detection of five new susceptibility loci not previously reported for CRC. The discovered loci were found to explain 10% of the overall risk of developing CRC. The strongest association was observed for rs10935945 in long non-coding RNA *LINC02006* (3q25.2). Three other SNPs were also located within genes (rs17575184 in *NEGR1*, rs11060839 in *PIWIL1*, rs12935896 in *BCAS3*), while one was intergenic (rs9927668 at 16p13.2). An expression quantitative trait locus (eQTL) bioinformatic analysis suggested that these polymorphisms may affect transcription factor binding sites. In conclusion, four of the identified variants were located within genes likely involved in tumor invasiveness and metastasis. Therefore, they could possibly be markers of poor prognosis in CRC patients.

Keywords: colorectal cancer; genome-wide association study; tumor progression; metastasis; long non-coding RNA; polymorphism

1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed malignant tumor, both around the world and in Poland [1]. It also represents the second and third leading cause of cancer-related deaths among men and women in the Polish population, respectively [2]. Genetic factors are thought to account for up to 35% of the variation in CRC risk [3,4]. Rare mutations with high penetration are responsible for less than 6% of cases of CRC [5,6]. In order to explain some of the remaining risk that contribute to CRC, numerous genome-wide association studies (GWAS) have been conducted for common low-penetrance variants. Currently, at least 100 independent susceptibility loci associated with CRC development at $p < 5 \times 10^{-8}$ have been identified, including over 50 new loci discovered in large-scale GWAS meta analyses in 2019 alone [7–12]. Despite these great efforts, less than 12% of familial relative risk [11] and less than 1% of the heritability of CRC [13] can be explained by the common variants identified by GWAS. Thus, most of the genetic factors contributing to the risk of CRC remain undetermined.

Given the significant population diversity in terms of genetic variation, and thus differences in allele frequencies and association strength, conducting analyses on different populations increases the chance of identifying general risk variants [14]. In addition, including ethnic or racial minorities can help to discover new loci or risk variants specific to the studied populations [15]. Our previous studies indicated that there are some benefits of studying relatively small but homogenous populations, such as the Polish population [16,17].

In several studies from recent years, we successfully implemented a novel approach for selecting GWAS-discovered single-nucleotide polymorphisms (SNPs) for further validation of their association with the disease [17–20]. This was shown to be effective particularly for GWAS with pooled DNA or with a small sample size, where limited study power hardly allows associations to be made at the standard genome-wide significance level ($p < 5 \times 10^{-8}$). In this approach, index SNPs for individual genotyping are selected based more on biological context than a purely statistical criterion, assuming that each associating SNP is usually not independent of neighboring variants. Such a method reduces the number of false-positive genome-wide associations and allows for the discovery of new associations [18]. Here, we adopted this method of selecting the index SNP in combination with pooled DNA sample GWAS for CRC susceptibility variants in the Polish population. This approach enabled the detection of five new susceptibility variants which have not yet been associated with CRC. Of these, four are intron variants of genes that are involved or very likely to be involved in the neoplastic process, especially tumor progression and metastasis. Interestingly, the strongest association was observed for the novel long non-coding RNA (lncRNA) variant, suggesting its role through interaction with transcription factors (TFs).

2. Materials and Methods

2.1. Ethics Statement

All patients and control subjects were Polish Caucasians recruited from two urban populations, Warsaw and Szczecin. The local ethics committee approved the study (Maria Skłodowska-Curie National Research Institute of Oncology, Warsaw, Poland, project ID: 37/2017/1/2021), and all subjects provided informed consent before they participated in the study. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

2.2. Patients

In total, the pooled DNA sample-based GWAS and verification stage with individual samples included 2013 individuals—465 with CRC and 1548 controls. The pooled-sample GWAS cohorts included 432 patients with CRC (168 females and 264 males; median age: 66 years; range: 20–91 years) and 672 control subjects (360 females and 312 males; median age: 55 years; range: 19–95 years). The demographic and clinical characteristics of all patients and controls are shown in Table 1.

Table 1. The demographic and clinical characteristics of patients and controls.

	CRC (N = 465)	Control (N = 1548)
Female N (%)	176 (38)	969 (63)
Male N (%)	289 (62)	579 (37)
Age (mean ± SD)	66 ± 11	55 ± 11
Age (median)	66	58
Age (min.–max.)	20–91	19–95
Tumor localization (%)		
rectum	173 (37.2)	
sigmoid	79 (17.0)	
sigmoid-rectum	72 (15.5)	
caecum	55 (11.8)	
ascendant	39 (8.4)	
other	47 (10.1)	
Tumor size (%)		
0	4 (0.9)	
1	40 (8.6)	
2	85 (18.3)	
3	277 (59.6)	
4	56 (12.0)	
Tis	3 (0.6)	
Node status (%)		
0	245 (52.7)	
1	126 (27.1)	
2	78 (16.8)	
3	9 (1.9)	
Nx	7 (1.5)	
Grade (%)		
1	27 (5.8)	
2	284 (61.1)	
3	44 (9.5)	
Gx	110 (23.6)	
Metastasis (%)		
	49 (10.5)	

CRC, colorectal cancer; N, number of subjects; SD, standard deviation; Tis, tumor in situ; Nx, indeterminate; Gx, indeterminate.

2.3. Genome-Wide Microarray Allelotyping

A pooled DNA sample-based GWAS was performed as described previously [21]. Genomic DNA was extracted from whole blood treated with EDTA using a QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), quantified using a Quant-iT™ PicoGreen dsDNA Kit (Invitrogen, Carlsbad, CA, USA), and visually checked for integrity on 1% agarose gel. Solely DNA samples that passed quality control tests (for purity, quantity, and integrity) were combined according to diagnosis and gender at equimolar concentrations to obtain 24-sample pools. A total of 18 DNA pools were prepared for the CRC group (seven for women and 11 for men) and 28 for controls (15 for women and 13 for men). Pooled DNA samples were adjusted to a final concentration of 50 ng/L in Tris-EDTA buffer (pH = 8) and analyzed individually on Illumina Infinium Omni2.5-Exome-8 v1.3 BeadChip microarrays by a commercial organization (Eurofins Genomics, Galten, Denmark). The datasets from GWAS are available from the Gene Expression Omnibus (GEO) database under accession number GSE156411.

2.4. Individual Genotyping

According to our previously described approach [17–20] for the verification of GWAS findings, loci were chosen that were represented by blocks of SNPs associated with CRC at the $p < 5 \times 10^{-3}$, for which the intervals between all pairs of adjacent SNPs were <30 kb.

From each of the independent loci, the most strongly associated SNP (at $p < 10^{-4}$) was selected as an index SNP for further verification via individual DNA sample genotyping, and stepwise forward logistic regression analysis. TaqMan SNP Genotyping Assays (Thermo Fisher Scientific, Waltham, MA, USA, a SensiMix™ II Probe Kit (Bioline Ltd., London, United Kingdom), and a 7900HT Real-Time PCR system (Thermo Fisher Scientific, Waltham, MA, USA) were used for individual genotyping in a 384-well format.

2.5. eQTL Analysis

Data collected in an online bioinformatics database by HaploReg, version 4.1 [22], was used for the analysis of identified CRC susceptibility genetic variants and the expression of quantitative trait loci (eQTL).

2.6. Survival Curves

Kaplan-Meier survival curves were prepared in Human Protein Atlas [23] (<http://www.proteinatlas.org>; accessed on 5 May 2021) using the default cut-off for differentiation between low and high gene expression and data from The Cancer Genome Atlas (TCGA) for the CRC patient cohort. *P*-values in the log-rank test were computed using default observation periods.

2.7. Statistical Analyses

2.7.1. Genome-Wide Allelotyping

First, the relative allele signal (RAS) for each SNP was calculated as $A/(A + B)$, where *A* and *B* were signal intensities for *A* and *B* alleles (as defined by Illumina). The RAS was used as an approximation of the allele ratio. Student's *t*-test (Welch variant) was used to compare allele ratios between groups. Due to a lack of the call-rate statistics for pooled samples, the quality was assessed via visual inspection of the first two principal components for outliers (Figure S1). One control and one CRC pool were removed. The calculated lambda value of 1.007, together with the quantile–quantile (Q–Q) plot of *p*-values (Figure S2), raised no concerns regarding the homogeneity of the final population. No probe filtering was performed. *P*-values were corrected for multiple hypothesis testing with the Holm algorithm. Manhattan plotting was performed using the qqman R package [24]. Probe names were mapped to a reference SNP ID using mapping files (InfiniumOmni2-5-8v1-5 and InfiniumOmniExpressExome-8v1-6) provided by Illumina. The impact of variants on the coding sequence and clinical significance were imported with the biomaRt package [25,26]. All computations were performed according to the R environment [27]. The study power calculations were performed using the epiR package [28], assuming the proportion of an allele in a reference group of 0.05–0.5 and an odds ratio (OR) of 1.2–2 (Table S1). The method assumed a confidence level of 95% in an unmatched case–control study.

2.7.2. Individual Genotyping

The Hardy–Weinberg equilibrium concordance of SNPs selected for verification was tested using the HardyWeinberg R package, version 1.6.8 [29], whereby no statistically significant deviations were observed. Differences in frequencies between groups were verified using the chi-squared test for alleles and the Cochran–Armitage test (implemented in R package DescTools, version 0.99.23) [30] for genotypes, with the exception of rs17575184, for which Fisher's exact test was chosen due to the small number of alternative homozygotes. The *p*-value significance threshold was adjusted for multiple comparisons with the Benjamini–Hochberg algorithm [31]. The OR and 95% confidence interval (CI) were estimated by normal approximation implemented in the EpiTools R package, version 0.5–10 [32]. OR values were given, with the more frequent allele or genotype taken as the reference.

2.7.3. Stepwise Forward Logistic Regression Analysis

Prediction analysis was performed by a stepwise forward logistic regression method, with the Akaike information criterion (AIC) used as the criterion for variable choice, using the step function of the R basic statistics package. The significant SNPs ($p < 0.05$) were ranked according to their AIC values, starting from a variant with the lowest AIC value, and sequentially introduced into the prediction model. Nagelkerke's pseudo- R^2 for each step was computed with the DescTools package, version 0.99.23 [30], in order to estimate the proportion of the overall risk of developing CRC. The area under the curve (AUC) value describing the accuracy of the prediction was computed using the pROC package, version 1.10 [33].

3. Results

3.1. Association Analyses

A pooled DNA sample-based GWAS in combination with a novel approach to SNP selection was applied in the search for new genetic variants associated with CRC in a Polish population. The 24-sample pools of DNA were used, obtained from 432 patients with CRC (18 pools) and 672 control subjects (28 pools). Seven independent loci were selected for further verification of their relationship to CRC development; neither showed statistically significant deviations from the Hardy–Weinberg equilibrium. Of these, six loci were represented by blocks of at least 11 SNPs associated with $p < 5 \times 10^{-3}$ at a distance of less than 30 kb from one another. One block (represented by rs17575184) consisted of seven SNPs, however three associated at $p < 10^{-4}$. From each selected locus, the most strongly associated index SNP was further verified via genotyping of individual DNA samples from both the CRC ($N = 465$) and control ($N = 1079$) groups.

As shown in Table 2, in verification analyses, five of the GWAS-selected SNPs exhibit significant differences in allele and genotype frequencies between the CRC and control groups after the Benjamini–Hochberg algorithm's adjustment for multiple testing ($p_{adj} < 0.05$). None of these associations have previously been reported for CRC. The strongest association was observed for rs10935945 in *LINC02006* at 3q25.2 ($p_{adj} = 1.26 \times 10^{-5}$ and 1.29×10^{-5} for allele and genotype frequencies, respectively). The next three SNPs revealed allelic associations at $p_{adj} \leq 3.92 \times 10^{-4}$. Apart from rs10935945, three other SNPs were located within gene regions (rs17575184 in *NEGR1* at 1p31.1, rs11060839 in *PIWIL1* at 12q24.33, and rs12935896 in *BCAS3* at 17q23.2), while one was at an intergenic location (16p13.2).

Additional association analyses, including stratifying the CRC patients cohort by tumor localization or different disease parameters (e.g., grading, metastasis), indicated that all five identified variants were significantly associated ($p_{adj} < 0.05$) with G2 and T3 CRCs (Table S2). Interestingly, some significant associations were observed even when the CRC subgroup was very small, e.g., rs11060839 with G3 ($N = 42$) or rs10935945 with T4 ($N = 56$) and N2 ($N = 78$).

The minor allele (MA) of two SNPs was associated with an increased risk of CRC development, while that of the remaining three SNPs showed a protective effect (Table 2). The effect size of all five susceptibility loci was relatively moderate ($OR \geq 1.45$ or ≤ 0.77), which is consistent with the estimated statistical power of our GWAS. Assuming an allele frequency of 0.3 to 0.5, a power ranging from 88% to 90% is needed to detect an effect size of $OR = 1.5$ (Table S1). The strongest effect was observed for rs17575184, located in the intron sequence of the *NEGR1* gene ($OR = 0.57$, 95% CI 0.42–0.76, $p_{adj} = 3.54 \times 10^{-4}$).

Table 2. The allelic and genotypic association of GWAS-selected, single-nucleotide polymorphisms (SNPs) with colorectal cancer.

dbSNP ID ^a	Region	MA	Allele Frequency (%)				Genotype Frequency (%)					
			MAF ^b	Control	CRC	OR (95% CI)	<i>p</i> _{adj} -Value	Genotype	Control	CRC	OR (95% CI)	<i>p</i> _{adj} -Value
rs17575184	1p31.1 <i>NEGR1</i> intron	A	0.088	232 (10.8)	60 (6.5)	0.57 (0.42–0.76)	3.54 × 10^{−4}	AA	11 (1.0)	1 (0.2)	0.22 (0.01–1.12)	7.91 × 10^{−4}
								AG	210 (19.6)	58 (12.5)	0.58 (0.42–0.79)	
								GG	852 (79.4)	406 (87.3)	-	
rs10935945	3q25.2 <i>LINC02006</i> intron	T	0.399	906 (42.2)	478 (51.5)	1.46 (1.25–1.70)	1.26 × 10^{−5}	TT	195 (18.2)	117 (25.2)	2.11 (1.54–2.90)	1.29 × 10^{−5}
								TC	516 (48.0)	244 (52.6)	1.66 (1.28–2.18)	
								CC	363 (33.8)	103 (22.2)	-	
rs10838094	11p15.4 <i>OR51B5</i> intron	A	0.378	445 (41.4)	422 (45.5)	1.18 (0.99–1.41)	8.03 × 10 ^{−2}	AA	97 (18.1)	92 (19.8)	1.34 (0.93–1.92)	8.12 × 10 ^{−2}
AG	251 (46.7)	238 (51.3)	1.34 (1.01–1.78)									
GG	189 (35.2)	134 (28.9)	-									
rs12424924	12p12.1 <i>PYROXD1</i> intron	A	0.194	223 (20.5)	165 (17.9)	0.85 (0.68–1.06)	0.147	AA	25 (4.6)	16 (3.5)	0.71 (0.37–1.36)	0.152
AG	173 (31.8)	133 (28.9)	0.86 (0.65–1.12)									
GG	346 (63.6)	311 (67.6)	-									
rs11060839	12q24.33 <i>PIWIL1</i> intron	A	0.169	332 (15.6)	194 (21.1)	1.45 (1.19–1.76)	3.92 × 10^{−4}	AA	27 (2.5)	21 (4.6)	2.06 (1.13–3.71)	5.70 × 10^{−4}
								AG	278 (26.1)	152 (33.0)	1.45 (1.14–1.84)	
								GG	760 (71.4)	287 (62.4)	-	
rs9927668	16p13.2 intergenic	C	0.290	840 (39.1)	285 (30.9)	0.70 (0.59–0.82)	5.04 × 10^{−5}	CC	179 (16.7)	46 (10.0)	0.48 (0.33–0.68)	8.25 × 10^{−5}
								CT	482 (44.9)	193 (41.9)	0.74 (0.59–0.94)	
								TT	412 (38.4)	222 (48.2)	-	
rs12935896	17q23.2 <i>BCAS3</i> intron	C	0.400	545 (25.4)	194 (20.9)	0.77 (0.64–0.93)	8.85 × 10^{−3}	CC	68 (6.4)	21 (4.5)	0.63 (0.37–1.03)	8.87 × 10^{−3}
								CT	409 (38.2)	152 (32.7)	0.76 (0.60–0.95)	
								TT	594 (55.5)	292 (62.8)	-	

Allelic frequencies of all studied SNPs were in Hardy–Weinberg equilibrium. Bold denotes significant association after Benjamini–Hochberg algorithm adjustment ($p_{adj} < 0.05$). CRC, colorectal cancer; MA, minor allele; MAF, MA frequency; OR, odds ratio; CI, confidence interval. ^{a/} SNP identifier based on NCBI SNP database (<http://www.ncbi.nlm.nih.gov/snp/>; accessed on 20 November 2020). ^{b/} MAF based on NCBI SNP database (<http://www.ncbi.nlm.nih.gov/snp/>; accessed on 20 November 2020).

3.2. Risk Prediction Modeling

To evaluate the contribution of individual SNPs to the risk of developing CRC, a stepwise forward logistic regression was performed with AIC minimization as a selection criterion. SNPs significant in the stepwise logistic regression were ranked according to their AIC value and sequentially introduced into the prediction model. Out of the seven SNPs selected for verification, only rs12424924 was excluded from further modeling ($p > 0.05$). According to the AIC estimates, the optimal model included six SNPs: rs9927668, rs10935945, rs17575184, rs12935896, rs11060839, and rs10838094 (Table 3). SNP rs9927668 emerged as the model with the lowest AIC value of all single-SNP models, and the addition of rs10935945 resulted in the largest AIC decrease. Both of these SNPs account for more than half of the risk of CRC development explained by the final model involving six SNPs. The sequential introduction of rs17575184, rs12935896, and rs11060839 moderately improved the parameters of the resulting models, while the inclusion of rs10838094 only marginally lowered the AIC. In total, the six SNPs included in the model were found to explain 10% of the overall risk of CRC development, as assessed using the Nagelkerke pseudo- R^2 statistic (Table 3). However, the overall accuracy expressed by an AUC value of 0.66 suggests the rather low predictability of the final model.

Table 3. The results of the stepwise selection for the logistic regression model.

dbSNP ID ^a	AIC ^b	AIC Change (%)	R^2 ^c	R^2 Change (%)
rs9927668	1309.45		0.028	
rs10935945	1294.45	15.0 (1.15)	0.054	0.026 (92)
rs17575184	1285.31	9.14 (0.74)	0.071	0.017 (33)
rs12935896	1279.74	5.57 (0.43)	0.084	0.013 (18)
rs11060839	1274.75	4.99 (0.39)	0.095	0.012 (14)
rs10838094	1274.26	0.49 (0.04)	0.101	0.006 (6)

Six significant SNPs ($p < 0.05$) ranked by Akaike information criterion (AIC) values were sequentially implemented into the model, starting with SNP rs9927668 with the lowest AIC value. All six SNPs were included in the final prediction model. ^a/ SNP identifier based on NCBI SNP database (<http://www.ncbi.nlm.nih.gov/SNP/>; accessed on 20 November 2020). ^b/ AIC value calculated after sequential implementation of the ranked SNPs. ^c/ Nagelkerke pseudo- R^2 value calculated after sequential implementation of the ranked SNPs.

3.3. eQTL Bioinformatic Analysis

A search of the HaploReg database [22,34] revealed that four of the five SNPs significantly associated with CRC risk potentially changed TF binding motifs, which may implicate the regulatory effect of a variant (Figure 1). Among others, the predicted binding motifs of D-box binding PAR BZIP transcription factor (DBP) and CCAAT/enhancer-binding protein gamma (CEBPG) TFs overlap with the position of rs10935945 and rs12935896 polymorphisms, respectively.

3.4. Survival Probability

The expression levels of *PIWIL1*, *NEGR1*, and *BCAS3* did not significantly affect the probability of survival when analyzed in the full cohort of CRC patients and controls ($p > 0.05$; Figure S3). However, after stratification by gender, high relative *PIWIL1* expression at diagnosis was associated with a significantly lower survival probability in the male cohort ($p = 0.02$, 46% vs. 77% 5-year survival probability; Figure S3B), while low expression of *NEGR1* was correlated with a significantly lower 5-year survival probability among women ($p = 0.049$, 54% vs. 65%; Figure S3D).

Position Weight Matrix ID	Ref	Alt	Delta	Polymorphism and Motif Sequence
DBP	-1.37	10.60	11.97	rs10935945 (LINC02006) Ref: ATGAAGTACATAGATTGAATTTCAAACAGTAAACCATCCTTACATTGCTATAATATACC Alt: ATGAAGTACATAGATTGAATTTCAAACAGCAAACCATCCTTACATTGCTATAATATACC AGCWHWH
CEBPG	0.51	12.48	11.97	rs12935896 (BCAS3) Ref: GCGGATGAGGAAACTGATGAGATCTGAAACGATTGCGCAATCTATCACAGCGAATAAG Alt: GCGGATGAGGAAACTGATGAGATCTGAAATGATTGCGCAATCTATCACAGCGAATAAG WTTYTKWMRTVRR
GATA_known11	13.01	12.59	-0.41	rs17575184 (NEGR1) Ref: CCTATGTTAGAACTCAAAGAAGTACCCTGTAATCTTGGTTCCAAAAGTGTGTTGTTGTA Alt: CCTATGTTAGAACTCAAAGAAGTACCCTATAATCTTGGTTCCAAAAGTGTGTTGTTGTA NYNVYAATCHBSMY
TATA_known5	13.86	13.58	-0.27	rs9927668 (intergenic) Ref: CAGAGGATCCAACCGACTCCCTTTTAAAGTTGTGCACCAGGGAACAGATGCTAAATGTG Alt: CAGAGGATCCAACCGACTCCCTTTTAAAGCTGTGCACCAGGGAACAGATGCTAAATGTG YTTTTAWRSN
Znf143_known1	8.58	11.39	2.80	KCDWKGMABMTGGRWVDKV

Figure 1. Possibly altered regulatory motifs. Based on HaploReg database (<https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>) [22]. Position weight matrix motifs in biological sequences [34]. Allele variants are indicated in bold. Ref, reference sequence; Alt, alternative sequence.

4. Discussion

An important issue in GWAS is the large number of false positive results, to some extent due to the structure of the studied population, but also due to certain methodological assumptions. Defining an appropriate p -value threshold for statistical significance appears to be critical [35,36]. Our previous study revealed that most associations selected solely on the basis of the arbitrarily established genome-wide significance level ($p < 5 \times 10^{-8}$) turned out to be false positives, whereas the inclusion of biological context in the SNP selection method (by taking into account the strong allele linkage disequilibrium (LG)) significantly reduced the number of false positives [18]. This approach also increases the chance of finding new associations in small-sized studies and GWAS based on pooled DNA samples when the ability to reach the standard genome-wide significance level is limited [17,19]. The adoption of this method of selecting the index SNP for verification in combination with pooled DNA sample-based GWAS enabled the detection of five new genetic variants associated with CRC development in the Polish population. None of these SNPs have previously been reported to be associated with CRC. It cannot be determined whether the finding of new susceptibility loci resulted from the adopted methodological approach or because these variants are more specific to our population. It should be noted, however, that in the group of our study subjects, both clearly higher and lower frequencies of the MA of the identified variants were observed compared to those reported in the NCBI SNP database (Table 2), e.g., SNPs rs9927668 (0.391 vs. 0.290) and rs12935896 (0.252 vs. 0.400), respectively.

The strongest effect on CRC susceptibility was observed for SNP rs17575184, located within an intron of the neuronal growth regulator 1 (*NEGR1*) gene (OR = 0.57; Table 2). Previous GWASs have indicated an association of the rs17575184 polymorphism with asthma in children ($p = 4 \times 10^{-3}$) [37], whereas other SNPs in *NEGR1* were implicated in body weight regulation [38,39] and dyslexia [40]. *NEGR1* is an extracellular adhesion protein that binds to cell membrane rafts, especially in the cell junction area, where it promotes cell-to-cell attachment and aggregation [41]. Given that adhesion properties are crucial in tumor cell migration and invasion during metastasis, *NEGR1* may play a role in

malignant transformation by regulating intercellular and cell-to-matrix interactions [42]. Accordingly, *NEGR1* was identified as a commonly downregulated gene in various types of human cancers, including CRC, suggesting its contribution to tumor suppression [41], while *NEGR1* overexpression reduced the tumorigenic properties of ovarian cancer cell line SKOV-3 cells [41]. When analyzed in the TCGA female CRC patients cohort, low expression of *NEGR1* was associated with the shorter 5-year survival rate (Figure S3D). *NEGR1* may play a role in regulating neurite outgrowth and neuronal arborization via direct interaction with receptor tyrosine kinase fibroblast growth factor receptor 2 (*FGFR2*) [43,44]. *FGFR2* silencing inhibited cell migration and invasion [45], and its overexpression negatively correlated with overall CRC patient survival [46]. Therefore, *NEGR1* may be functionally related to CRC, although the variant identified by GWAS is unlikely to be directly causal.

Similar to *NEGR1*, the breast-carcinoma-amplified sequence 3 (*BCAS3*) gene encoded a protein involved in cell adhesion and migration processes. The rs12935896, located in the intron sequence of *BCAS3*, was also associated with decreased CRC susceptibility (OR = 0.77; Table 2). *BCAS3* polymorphisms were previously associated via GWAS with gout [47], traits of kidney disease [48], and coronary artery disease [49]. Its misexpression was found in various types of cancer [50,51], and was implicated in tumor progression to a higher grade of malignancy [52]. *BCAS3* is a cytoskeletal WD repeat domain-containing protein essential for angiogenesis, both during the developmental process and in tumor metastasis [51,53]. By activating and recruiting cell division cycle 42 (*CDC42*) Rho-GTPase [54] and facilitating crosstalk between cytoskeleton elements, *BCAS3* regulates cell polarity and focal adhesion assembly [55].

SNP rs11060839 located in the intronic sequence of PIWI-like, RNA-mediated gene silencing 1 (*PIWIL1*) was associated with an increased risk of developing CRC (OR = 1.45; Table 2). *PIWIL1* is a member of the PIWI-like family of Argonaute proteins, commonly associated with stem cell differentiation and self-renewal, RNA silencing, and the regulation of gene expression, whose activity is mediated by interactions with a specific class of small non-coding RNAs, referred to as PIWI-interacting RNAs (piRNAs) [56,57]. Both *PIWIL1* and piRNAs are overexpressed in CRC [58], and an upward trend was observed in *PIWIL1* expression levels during the colon adenoma–carcinoma sequence [59].

In patients with CRC, *PIWIL1* expression levels were closely related to the degree of tumor differentiation, TNM stage, the occurrence of lymph node invasion, and distant metastasis [59–61], suggesting that increased *PIWIL1* expression may promote tumor invasion. Moreover, patients with *PIWIL1* overexpression exhibited worse overall survival and disease-free survival, especially in the case of CRC at early stages or without lymph node invasion, showing the potential prognostic value of the *PIWIL1* expression status [61–63]. Accordingly, high expression levels of *PIWIL1* were associated with significantly lower 5-year survival probability when analyzed in the TCGA male CRC cohort (Figure S3B). Recently, a functional analysis of transcripts interacting with the *PIWIL1*–piRNA complex in the CRC COLO 205 cell line suggested that this complex may be directly involved in the activity regulation of key components of signal transduction cascades that are frequently dysregulated in CRC progression, including tumor suppressors and genes involved in the control of cell proliferation and differentiation, such as *IGF1R*, *JUN*, and *ERBB3* [64].

The strongest association reported in the current study was observed for rs10935945 in the lncRNA coding gene *LINC02006* at 3q25.2, which was associated with increased CRC risk. Although incapable of encoding proteins, lncRNAs play critical roles in the regulation of various cellular processes, such as cell growth, proliferation, apoptosis, and cancer progression [65]. lncRNAs can regulate gene expression, mainly at the post-transcriptional level, via various modes of direct action or as miRNA sponges or endogenous competitors, thus reducing their regulatory effect on target mRNAs [66]. The aberrant expression of lncRNAs, exemplified by *lncRNA H19* and *lncRNA 91H*, an antisense gene of *H19*, has been implicated in the tumorigenesis and metastasis of different types of cancer, including CRC, where it is associated with a poor prognosis and a high risk of tumor metastasis [67,68]. Both *LINC01354* and *lncRNA CASC11* are upregulated in CRC and contribute to the pro-

liferation, invasion, and metastasis of CRC via activation of the Wnt/ β -catenin signaling pathway [69,70]. Additionally, *LINC01123* was upregulated in CRC tumors and cells, and its expression positively correlated with the vascular endothelial growth factor A (*VEGFA*) expression and the binding of miR-34c-5p, sponged by *LINC01123* [71]. The silencing of *UNC5B-AS1*, highly expressed in CRC tissues, repressed cancer growth and metastasis, most likely by increasing miR-622 expression and suppression of the AMP-activated protein kinase (AMPK) and the phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) signaling pathways [72]. On the other hand, overexpression of *lncRNA TUSC7* reduces cell migration and invasion in CRC by sponging miR-211 [73].

Several genetic variants located in lncRNA genes influence the risk of CRC development; polymorphisms in *lncRNA HOTTIP*, rs145204276 and rs55829688 in *lncRNA GAS5*, rs2839698 in *lncRNA H19*, rs2632159 in *lncRNA PCAT1*, rs2147578 in *lnc-LAMC2-1:1*, and rs664589 in *MALAT1* were associated with a significantly increased CRC risk [67], while rs13252298 and rs1456315 in *lncRNA PRNCR1* and rs1194338 in *MALAT1* had protective effects on CRC [67,74]. It has been suggested that the rs664589 G allele alters the binding of *MALAT1* to miR-194-5p, resulting in an increased expression of *MALAT1* and enhanced CRC development and metastasis [75], while the rs2147578 in *lnc-LAMC2-1:1* affects the sponging of miR-128-3p, which correlates with higher expression of the *LAMC2* oncogene in CRC [76]. The rs55829688 in the *lncRNA GAS5* promoter region exerts its regulatory effect by affecting the binding affinity of TF YY1 to the *GAS5* promoter and downregulating *GAS5* expression [77]. Additionally, the *lncRNA PCAT1* rs2632159 may impact the risk of CRC by modulating the binding of EBF, LUN-1, and TCF12 [78], which could stimulate *PCAT1* expression, thus increasing its oncogenic function. Similarly, an eQTL bioinformatic analysis showed that the rs10935945 T variant of *LINC02006* identified in this study could influence binding with the TF DBP, a member of the PAR leucine zipper TF family (Figure 1), possibly increasing the risk of CRC development.

Based on the above-mentioned functional relations to CRC, it can be speculated that the gene variants identified in this study may be associated with metastatic and invasive CRC. Accordingly, after stratification of the CRC patient cohort by different disease parameters, all five identified variants were significantly associated with rather advanced tumors (T3; Table S2). Moreover, when comparing the CRC subgroup with the metastases to the control group, associations of three variants (rs11060839, rs9927668, and rs12935896) were observed, although they did not remain significant after statistical adjustment. However, it should be borne in mind that the lack of certain associations could be due to the small size of the analyzed subgroups, and all these observations need to be validated in a bigger, independent study. Our study included CRC patients with relatively advanced neoplastic disease, as indicated by the clinical characteristics of the patients (Table 1). Apart from the differences resulting from the use of technologically different platforms, this may also be the reason why the associations of known GWAS variants did not reach statistical significance. Moreover, the specific genetic architecture of the Polish population may also be important, which is consistent with the results of the recent replication study in the Basque population [79].

5. Conclusions

In this study, five new susceptibility variants associated with CRC development were revealed by the pooled DNA sample GWAS in a Polish population. Among them, four are intron variants of genes encoding proteins that are likely involved in the neoplastic process, especially tumor invasiveness and metastasis, and therefore could possibly be markers of poor prognosis in CRC patients. In total, discovered loci were found to account for 10% of the variation in the risk of developing CRC. While the prediction accuracy of the built model was rather low, the newly identified variants can significantly improve the cumulative risk assessment of CRC based on common susceptibility variants.

In line with the growing body of data suggesting that SNPs in lncRNAs can influence CRC risk, the novel lncRNA variant *LINC02006* was shown to express the strongest

association with CRC development, possibly by affecting the DBP TK binding site and deregulating downstream pathways. Further understanding of lncRNA functions in cancer progression could improve CRC prediction and diagnosis.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/biology10060465/s1>: Figure S1: A plot of the first two principal components for the relative allele signal (RAS) values of all analyzed microarray samples. Figure S2: A quantile–quantile (Q–Q) plot of the 2,561,761 *t* statistics from the whole-exome microarray analysis of 18 CRC and 28 control pooled DNA samples. Figure S3: The Kaplan–Meier curves for the association of the expression levels of identified genes with survival probability in colorectal cancer patients, according to the *Human Protein Atlas* database. Table S1: The power of the study calculated for a given minor allele frequency (first column) and odds ratio (second row). Table S2: The significant allelic association of GWAS-selected SNPs with colorectal cancer parameters.

Author Contributions: Conceptualization, E.E.H. and J.O.; investigation, A.K., M.P., A.B., N.Z.-L., F.A., and J.K.; validation, A.K., M.P., A.B., M.K., and E.E.H.; formal analysis, M.K., K.G., A.K., N.Z.-L., and E.E.H.; resources, T.O. and Ł.Z.; data curation, K.G., M.K., A.K., and E.E.H.; writing—original draft preparation, E.E.H. and M.K.; writing—review and editing, N.Z.-L., M.K., K.G., A.K., and E.E.H.; visualization, K.G. and E.E.H.; supervision, E.E.H. and J.O.; project administration, E.E.H. and J.O.; funding acquisition, J.O. and E.E.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by internal grant No. SN/GW01/2017 from Maria Skłodowska-Curie National Research Institute of Oncology, Warsaw, Poland (www.pib-nio.pl). E.E.H. was supported by grant No. 501-1-009-12-20 from the Centre of Postgraduate Medical Education, Warsaw, Poland (www.cmkp.edu.pl).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Maria Skłodowska-Curie National Research Institute of Oncology, Warsaw, Poland, (protocol 37/2017/1/2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available in this article. The datasets from GWAS are available at the Gene Expression Omnibus (GEO) database under accession number GSE156411.

Acknowledgments: The authors thank A. Paziewska (Centre of Postgraduate Medical Education, Warsaw, Poland) for contributions to project administration, participation in investigations, and scientific advising. We are grateful to M. Dąbrowska (Maria Skłodowska-Curie National Research Institute of Oncology, Warsaw, Poland) for creating the programs necessary to conduct some of the analyzes. We also acknowledge the contributions of P. Cybula (Centre of Postgraduate Medical Education, Warsaw, Poland) in sample preparation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2014**, *136*, E359–E386. [[CrossRef](#)]
2. Religioni, U. Cancer incidence and mortality in Poland. *Clin. Epidemiol. Glob. Health* **2020**, *8*, 329–334. [[CrossRef](#)]
3. Lichtenstein, P.; Holm, N.V.; Verkasalo, P.K.; Iliadou, A.; Kaprio, J.; Koskenvuo, M.; Pukkala, E.; Skytthe, A.; Hemminki, K. Environmental and heritable factors in the causation of cancer—Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **2000**, *343*, 78–85. [[CrossRef](#)] [[PubMed](#)]
4. Mucci, L.A.; Hjelmborg, J.; Harris, J.R.; Czene, K.; Havelick, D.J.; Scheike, T.; Graff, R.E.; Holst, K.; Möller, S.; Unger, R.H.; et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA* **2016**, *315*, 68–76. [[CrossRef](#)]
5. De La Chapelle, A. Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **2004**, *4*, 769–780. [[CrossRef](#)] [[PubMed](#)]
6. Peters, U.; Bien, S.; Zubair, N. Genetic architecture of colorectal cancer. *Gut* **2015**, *64*, 1623–1636. [[CrossRef](#)] [[PubMed](#)]
7. Huyghe, J.R.; Bien, S.A.; Harrison, T.A.; Kang, H.M.; Chen, S.; Schmit, S.L.; Conti, D.V.; Qu, C.; Jeon, J.; Edlund, C.K.; et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **2018**, *51*, 76–87. [[CrossRef](#)]

8. Law, P.J.; Timofeeva, M.; Fernandez-Rozadilla, C.; Broderick, P.; Studd, J.; Fernandez-Tajes, J.; Farrington, S.; Svinti, V.; Palles, C.; Orlando, G.; et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **2019**, *10*, 1–15. [[CrossRef](#)]
9. Lu, Y.; Kweon, S.-S.; Tanikawa, C.; Jia, W.-H.; Xiang, Y.-B.; Cai, Q.; Zeng, C.; Schmit, S.L.; Shin, A.; Matsuo, K.; et al. Large-scale genome-wide association study of east Asians identifies loci associated with risk for colorectal cancer. *Gastroenterology* **2019**, *156*, 1455–1466. [[CrossRef](#)]
10. MacArthur, J.; Bowler, E.; Cerezo, M.; Gil, L.; Hall, P.; Hastings, E.; Junkins, H.; McMahon, A.; Milano, A.; Morales, J.; et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **2016**, *45*, D896–D901. [[CrossRef](#)]
11. Schmit, S.L.; Edlund, C.K.; Schumacher, F.R.; Gong, J.; Harrison, T.A.; Huyghe, J.R.; Qu, C.; Melas, M.; Berg, D.J.V.D.; Wang, H.; et al. Novel common genetic susceptibility loci for colorectal cancer. *J. Natl. Cancer Inst.* **2018**, *111*, 146–157. [[CrossRef](#)] [[PubMed](#)]
12. Zeng, C.; Matsuda, K.; Jia, W.-H.; Chang, J.; Kweon, S.-S.; Xiang, Y.-B.; Shin, A.; Jee, S.H.; Kim, D.-H.; Zhang, B.; et al. Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology* **2016**, *150*, 1633–1645. [[CrossRef](#)] [[PubMed](#)]
13. Jiao, S.; Peters, U.; Berndt, S.; Brenner, H.; Butterbach, K.; Caan, B.; Carlson, C.S.; Chan, A.T.; Chang-Claude, J.; Chanock, S.; et al. Estimating the heritability of colorectal cancer. *Hum. Mol. Genet.* **2014**, *23*, 3898–3905. [[CrossRef](#)] [[PubMed](#)]
14. Jostins, L.; Barrett, J.C. Genetic risk prediction in complex disease. *Hum. Mol. Genet.* **2011**, *20*, R182–R188. [[CrossRef](#)] [[PubMed](#)]
15. Wang, H.; Schmit, S.L.; Haiman, C.A.; Keku, T.O.; Kato, I.; Palmer, J.R.; Berg, D.V.D.; Wilkens, L.R.; Burnett, T.; Conti, D.V.; et al. Novel colon cancer susceptibility variants identified from a genome-wide association study in African Americans. *Int. J. Cancer* **2017**, *140*, 2728–2733. [[CrossRef](#)]
16. Ledwoń, J.K.; Hennig, E.; Maryan, N.; Goryca, K.; Nowakowska, D.; Niwińska, A.; Ostrowski, J. Common low-penetrance risk variants associated with breast cancer in Polish women. *BMC Cancer* **2013**, *13*, 510. [[CrossRef](#)]
17. Zagajewska, K.; Piątkowska, M.; Goryca, K.; Bałabas, A.; Kluska, A.; Paziewska, A.; Pośpiech, E.; Grabska-Liberek, I.; Hennig, E.E. GWAS links variants in neuronal development and actin remodeling related loci with pseudoexfoliation syndrome without glaucoma. *Exp. Eye Res.* **2018**, *168*, 138–148. [[CrossRef](#)]
18. Ostrowski, J.; Paziewska, A.; Lazowska, I.; Ambrozkiwicz, F.; Goryca, K.; Kulecka, M.; Rawa, T.; Karczmarski, J.; Dabrowska, M.; Zeber-Lubecka, N.; et al. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Sci. Rep.* **2016**, *6*, 39831. [[CrossRef](#)]
19. Paziewska, A.; Habor, A.; Rogowska, A.; Zych, W.; Goryca, K.; Karczmarski, J.; Dabrowska, M.; Ambrozkiwicz, F.; Walewska-Zielecka, B.; Krawczyk, M.; et al. A novel approach to genome-wide association analysis identifies genetic associations with primary biliary cholangitis and primary sclerosing cholangitis in Polish patients. *BMC Med. Genom.* **2017**, *10*, 2. [[CrossRef](#)] [[PubMed](#)]
20. Hennig, E.E.; Piątkowska, M.; Goryca, K.; Pośpiech, E.; Paziewska, A.; Karczmarski, J.; Kluska, A.; Brewczyńska, E.; Ostrowski, J. Non-CYP2D6 variants selected by a GWAS improve the prediction of impaired tamoxifen metabolism in patients with breast cancer. *J. Clin. Med.* **2019**, *8*, 1087. [[CrossRef](#)]
21. Gaj, P.; Maryan, N.; Hennig, E.E.; Ledwon, J.K.; Paziewska, A.; Majewska, A.; Karczmarski, J.; Nesteruk, M.; Wolski, J.; Antoniewicz, A.A.; et al. Pooled sample-based GWAS: A cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population. *PLoS ONE* **2012**, *7*, e35307. [[CrossRef](#)] [[PubMed](#)]
22. HaploReg v4.1. Available online: <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> (accessed on 21 November 2020).
23. Uhlen, M.; Zhang, C.; Lee, S.; Sjödstedt, E.; Fagerberg, L.; Bidkhor, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507. [[CrossRef](#)] [[PubMed](#)]
24. Turner, S.D. qqman: An R package for visualizing GWAS results using Q–Q and manhattan plots. *J. Open Source Softw.* **2018**, *3*. [[CrossRef](#)]
25. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [[CrossRef](#)]
26. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [[CrossRef](#)]
27. R Development Core Team. *A Language and Environment for Statistical Computing: Reference Index*; R Foundation for Statistical Computing: Vienna, Austria, 2010; ISBN 978-3-900051-07-5.
28. Stevenson, M.; Nunes, T.; Heuer, C.; Marshall, J.; Sanchez, J.; Thornton, R.; Reiczigel, J. EpiR: Tools for the Analysis of Epidemiological Data. Available online: <https://cran.r-project.org/web/packages/epiR/index.html> (accessed on 21 November 2020).
29. Graffelman, J. Exploring diallelic genetic markers: The Hardy Weinberg package. *J. Stat. Softw.* **2015**, *64*, 1–23. [[CrossRef](#)]
30. Signorell, A.; Aho, K.; Alfons, A.; Anderegg, N.; Aragon, T.; Arachchige, C. DescTools: Tools for Descriptive Statistics. Available online: <https://CRAN.R-project.org/package=DescTools> (accessed on 20 November 2020).
31. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1995**, *57*, 289–300. [[CrossRef](#)]
32. Aragon, T.; Fay, M.; Wollschlaeger, D.; Omidpanah, A. Epitools: Epidemiology Tools. Available online: <https://CRAN.R-project.org/package=epitools> (accessed on 20 November 2020).

33. Robin, X.A.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Muller, M.J. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
34. Kheradpour, P.; Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **2013**, *42*, 2976–2987. [[CrossRef](#)]
35. Panagiotou, O.A.; Ioannidis, J.P.A.; Project, F.T.G.-W.S. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int. J. Epidemiol.* **2011**, *41*, 273–286. [[CrossRef](#)]
36. Kaler, A.S.; Purcell, L.C. Estimation of a significance threshold for genome-wide association studies. *BMC Genom.* **2019**, *20*, 1–8. [[CrossRef](#)]
37. Melén, E.; Himes, B.E.; Brehm, J.M.; Boutaoui, N.; Klanderma, B.J.; Sylvia, J.S.; Lasky-Su, J. Analyses of shared genetic factors between asthma and obesity in children. *J. Allergy Clin. Immunol.* **2010**, *126*, 631–637. [[CrossRef](#)]
38. Thorleifsson, G.; Walters, G.B.; Gudbjartsson, D.F.; Steinthorsdottir, V.; Sulem, P.; Helgadóttir, A.; Stykarsdóttir, U.; Gretarsdóttir, S.; Thorlacius, S.; Jonsdóttir, I.; et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **2008**, *41*, 18–24. [[CrossRef](#)]
39. The GIANT consortium. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **2008**, *41*, 25–34. [[CrossRef](#)]
40. Veerappa, A.M.; Saldanha, M.; Padakannaya, P.; Ramachandra, N.B. Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spinal plasticity. *J. Hum. Genet.* **2013**, *58*, 539–547. [[CrossRef](#)]
41. Kim, H.; Hwang, J.-S.; Lee, B.; Hong, J.; Lee, S. Newly identified cancer-associated role of human neuronal growth regulator 1 (NEGR1). *J. Cancer* **2014**, *5*, 598–608. [[CrossRef](#)]
42. Okegawa, T.; Pong, R.-C.; Li, Y.; Hsieh, J.-T. The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochim. Pol.* **2004**, *51*, 445–457. [[CrossRef](#)] [[PubMed](#)]
43. Epischedda, F.; Epiccoli, G. The IgLON family member negr1 promotes neuronal arborization acting as soluble factor via FGFR2. *Front. Mol. Neurosci.* **2016**, *8*, 89. [[CrossRef](#)]
44. Szczurkowska, J.; Pischedda, F.; Pinto, B.; Managò, F.; Haas, C.A.; Summa, M.; Bertorelli, R.; Papaleo, F.; Schäfer, M.K.; Piccoli, G.; et al. NEGR1 and FGFR2 cooperatively regulate cortical development and core behaviours related to autism disorders in mice. *Brain* **2018**, *141*, 2772–2794. [[CrossRef](#)]
45. Matsuda, Y.; Ishiwata, T.; Yamahatsu, K.; Kawahara, K.; Hagio, M.; Peng, W.-X.; Yamamoto, T.; Nakazawa, N.; Seya, T.; Ohaki, Y.; et al. Overexpressed fibroblast growth factor receptor 2 in the invasive front of colorectal cancer: A potential therapeutic target in colorectal cancer. *Cancer Lett.* **2011**, *309*, 209–219. [[CrossRef](#)] [[PubMed](#)]
46. Li, P.; Huang, T.; Zou, Q.; Liu, D.; Wang, Y.; Tan, X.; Wei, Y.; Qiu, H. FGFR2 promotes expression of PD-L1 in colorectal cancer via the JAK/STAT3 signaling pathway. *J. Immunol.* **2019**, *202*, 3065–3075. [[CrossRef](#)] [[PubMed](#)]
47. Sakiyama, M.; Matsuo, H.; Nakaoka, H.; Kawamura, Y.; Kawaguchi, M.; Higashino, T.; Nakayama, A.; Akashi, A.; Ueyama, J.; Kondo, T.; et al. Common variant of BCAS3 is associated with gout risk in Japanese population: The first replication study after gout GWAS in Han Chinese. *BMC Med. Genet.* **2018**, *19*, 96. [[CrossRef](#)]
48. Lee, J.; Lee, Y.; Park, B.; Won, S.; Han, J.S.; Heo, N.J. Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS ONE* **2018**, *13*, e0194044. [[CrossRef](#)]
49. Xiangfeng the CARDIoGRAMplusC4D Consortium. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **2015**, *47*, 1121–1130. [[CrossRef](#)] [[PubMed](#)]
50. Bärlund, M.; Monni, O.; Kononen, J.; Cornelison, R.; Torhorst, J.; Sauter, G.; Kallioniemi, O.-P.; Kallioniemi, A. Multiple genes at 17q23 undergo amplification and overexpression in breast cancer. *Cancer Res.* **2000**, *60*, 5340–5344. [[PubMed](#)]
51. Siva, K.; Venu, P.; Mahadevan, A.; Shankar, S.K.; Inamdar, M.S. Human BCAS3 expression in embryonic stem cells and vascular precursors suggests a role in human embryogenesis and tumor angiogenesis. *PLoS ONE* **2007**, *2*, e1202. [[CrossRef](#)]
52. Gururaj, A.E.; Singh, R.R.; Rayala, S.K.; Holm, C.; Hollander, P.D.; Zhang, H.; Balasenthil, S.; Talukder, A.H.; Landberg, G.; Kumar, R. MTA1, a transcriptional activator of breast cancer amplified sequence 3. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6670–6675. [[CrossRef](#)] [[PubMed](#)]
53. Shetty, R.; Joshi, D.; Jain, M.; Vasudevan, M.; Paul, J.C.; Bhat, G.; Banerjee, P.; Abe, T.; Kiyonari, H.; VijayRaghavan, K.; et al. Rudhira/BCAS3 is essential for mouse development and cardiovascular patterning. *Sci. Rep.* **2018**, *8*, 5632. [[CrossRef](#)] [[PubMed](#)]
54. Jain, M.; Bhat, G.P.; VijayRaghavan, K.; Inamdar, M.S. Rudhira/BCAS3 is a cytoskeletal protein that controls Cdc42 activation and directional cell migration during angiogenesis. *Exp. Cell Res.* **2012**, *318*, 753–767. [[CrossRef](#)]
55. Joshi, D.; Inamdar, M.S. Rudhira/BCAS3 couples microtubules and intermediate filaments to promote cell migration for angiogenic remodeling. *Mol. Biol. Cell* **2019**, *30*, 1437–1450. [[CrossRef](#)]
56. Han, Y.-N.; Li, Y.; Xia, S.-Q.; Zhang, Y.-Y.; Zheng, J.-H.; Li, W. PIWI proteins and PIWI-interacting RNA: Emerging roles in cancer. *Cell. Physiol. Biochem.* **2017**, *44*, 1–20. [[CrossRef](#)]
57. Rojas-Ríos, P.; Simonelig, M. piRNAs and PIWI proteins: Regulators of gene expression in development and stem cells. *Development* **2018**, *145*, dev161786. [[CrossRef](#)]
58. Litwin, M.; Dubis, J.; Arczyńska, K.; Piotrowska, A.; Frydlewicz, A.; Karczewski, M.; Dzięgiel, P.; Witkiewicz, W. Correlation of HIWI and HILI Expression with cancer stem cell markers in colorectal cancer. *Anticancer. Res.* **2015**, *35*, 3317–3324. [[PubMed](#)]
59. Wang, H.; Chen, B.; Cao, X.; Wang, J.; Hu, X.; Mu, X.; Chen, X. The clinical significances of the abnormal expressions of Piwil1 and Piwil2 in colonic adenoma and adenocarcinoma. *OncoTargets Ther.* **2015**, *8*, 1259–1264. [[CrossRef](#)] [[PubMed](#)]

60. Raeisossadati, R.; Abbaszadegan, M.R.; Moghbeli, M.; Tavassoli, A.; Kihara, A.H.; Forghanifard, M.M. Aberrant expression of DPPA2 and HIWI genes in colorectal cancer and their impacts on poor prognosis. *Tumor Biol.* **2014**, *35*, 5299–5305. [[CrossRef](#)] [[PubMed](#)]
61. Sun, R.; Gao, C.-L.; Li, D.-H.; Li, B.-J.; Ding, Y.-H. Expression status of PIWIL1 as a prognostic marker of colorectal cancer. *Dis. Markers* **2017**, *2017*, 1–7. [[CrossRef](#)]
62. Zeng, Y.; Qu, L.-K.; Meng, L.; Liu, C.-Y.; Dong, B.; Xing, X.-F.; Wu, J.; Shou, C.-C. HIWI expression profile in cancer cells and its prognostic value for patients with colorectal cancer. *Chin. Med. J.* **2011**, *124*, 2144–2149. [[PubMed](#)]
63. Liu, C.; Qu, L.; Xing, X.; Ren, T.; Zeng, Y.; Jiang, B.; Meng, L.; Wu, J.; Shou, C.; Dong, B. Combined phenotype of 4 markers improves prognostic value of patients with colon cancer. *Am. J. Med. Sci.* **2012**, *343*, 295–302. [[CrossRef](#)] [[PubMed](#)]
64. Sellitto, A.; Geles, K.; D'Agostino, Y.; Conte, M.; Alexandrova, E.; Rocco, D.; Nassa, G.; Giurato, G.; Tarallo, R.; Weisz, A.; et al. Molecular and Functional Characterization of the Somatic PIWIL1/piRNA Pathway in Colorectal Cancer Cells. *Cells* **2019**, *8*, 1390. [[CrossRef](#)]
65. Hu, X.; Sood, A.K.; Dang, C.V.; Zhang, L. The role of long noncoding RNAs in cancer: The dark matter matters. *Curr. Opin. Genet. Dev.* **2018**, *48*, 8–15. [[CrossRef](#)]
66. Sun, B.; Liu, C.; Zhang, L.; Luo, G.; Liang, S. Research progress on the interactions between long non-coding RNAs and microRNAs in human cancer (Review). *Oncol. Lett.* **2019**, *19*, 595–605. [[CrossRef](#)] [[PubMed](#)]
67. Poursheikhani, A.; Abbaszadegan, M.R.; Kerachian, M.A. Mechanisms of long non-coding RNA function in colorectal cancer tumorigenesis. *Asia Pacific J. Clin. Oncol.* **2020**, *17*, 7–23. [[CrossRef](#)]
68. Deng, Q.; He, B.; Gao, T.; Pan, Y.; Sun, H.; Xu, Y.; Li, R.; Ying, H.; Wang, F.; Liu, X.; et al. Up-regulation of 91H promotes tumor metastasis and predicts poor prognosis for patients with colorectal cancer. *PLoS ONE* **2014**, *9*, e103022. [[CrossRef](#)] [[PubMed](#)]
69. Li, J.; He, M.; Xu, W.; Huang, S. LINC01354 interacting with hnRNP-D contributes to the proliferation and metastasis in colorectal cancer through activating Wnt/ β -catenin signaling pathway. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 1–15. [[CrossRef](#)] [[PubMed](#)]
70. Zhang, Z.; Zhou, C.; Chang, Y.; Zhang, Z.; Hu, Y.; Zhang, F.; Lu, Y.; Zheng, L.; Zhang, W.; Li, X.; et al. Long non-coding RNA CASC11 interacts with hnRNP-K and activates the WNT/ β -catenin pathway to promote growth and metastasis in colorectal cancer. *Cancer Lett.* **2016**, *376*, 62–73. [[CrossRef](#)] [[PubMed](#)]
71. Ye, S.; Sun, B.; Wu, W.; Yu, C.; Tian, T.; Lian, Z.; Liang, Q.; Zhou, Y. LINC01123 facilitates proliferation, invasion and chemoresistance of colon cancer cells. *Biosci. Rep.* **2020**, *40*. [[CrossRef](#)] [[PubMed](#)]
72. Zhang, Y.; Li, Z.; Lan, Z. Silencing UNC5B antisense lncRNA 1 represses growth and metastasis of human Colon cancer cells via raising miR-622. *Artif. Cells Nanomed. Biotechnol.* **2019**, *48*, 60–67. [[CrossRef](#)]
73. Xu, J.; Zhang, R.; Zhao, J. The novel long noncoding RNA TUSC7 inhibits proliferation by sponging MiR-211 in colorectal cancer. *Cell. Physiol. Biochem.* **2017**, *41*, 635–644. [[CrossRef](#)]
74. Li, L.; Sun, R.; Liang, Y.; Pan, X.; Li, Z.; Bai, P.; Zeng, X.; Zhang, D.; Zhang, L.; Gao, L. Association between polymorphisms in long non-coding RNA PRNCR1 in 8q24 and risk of colorectal cancer. *J. Exp. Clin. Cancer Res.* **2013**, *32*, 104. [[CrossRef](#)]
75. Wu, S.; Sun, H.; Wang, Y.; Yang, X.; Meng, Q.; Yang, H.; Zhu, H.; Tang, W.; Li, X.; Aschner, M.; et al. MALAT1 rs664589 polymorphism inhibits binding to miR-194-5p, contributing to colorectal cancer risk, growth, and metastasis. *Cancer Res.* **2019**, *79*, 5432–5441. [[CrossRef](#)]
76. Gong, J.; Tian, J.; Lou, J.; Ke, J.; Li, L.; Li, J.; Yang, Y.; Gong, Y.; Zhu, Y.; Zhang, Y.; et al. A functional polymorphism in lnc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. *Carcinogenesis* **2016**, *37*, 443–451. [[CrossRef](#)]
77. Wang, Y.; Wu, S.; Yang, X.; Li, X.; Chen, R. Association between polymorphism in the promoter region of lncRNA GAS5 and the risk of colorectal cancer. *Biosci. Rep.* **2019**, *39*. [[CrossRef](#)] [[PubMed](#)]
78. Yang, M.-L.; Huang, Z.; Wu, L.-N.; Wu, R.; Ding, H.-X.; Wang, B.-G. lncRNA-PCAT1 rs2632159 polymorphism could be a biomarker for colorectal cancer susceptibility. *Biosci. Rep.* **2019**, *39*. [[CrossRef](#)] [[PubMed](#)]
79. Alegria-Lertxundi, I.; Aguirre, C.; Bujanda, L.; Fernández, F.J.; Polo, F.; Ordovás, J.M.; Etxezarraga, M.C.; Zabalza, I.; Larzabal, M.; Portillo, I.; et al. Single nucleotide polymorphisms associated with susceptibility for development of colorectal cancer: Case-control study in a Basque population. *PLoS ONE* **2019**, *14*, e0225779. [[CrossRef](#)]