



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Meta-Analysis of Mismatch Repair Polymorphisms within the Cogent Consortium for Colorectal Cancer Susceptibility

### Citation for published version:

Members of the EPICOLON Consortium (Gastrointestinal Oncology Group of the Spanish Gastroenterological Association):, Picelli, S, Lorenzo Bermejo, J, Chang-Claude, J, Hoffmeister, M, Fernández-Rozadilla, C, Carracedo, A, Castells, A, Castellví-Bel, S, Naccarati, A, Pardini, B, Vodickova, L, Müller, H, Talseth-Palmer, BA, Stibbard, G, Peterlongo, P, Nici, C, Veneroni, S, Li, L, Casey, G, Tenesa, A, Farrington, SM, Tomlinson, I, Moreno, V, van Wezel, T, Wijnen, J, Dunlop, M, Radice, P, Scott, RJ, Vodicka, P, Ruiz-Ponte, C, Brenner, H, Buch, S, Völzke, H, Hampe, J, Schafmayer, C & Lindblom, A 2013, 'Meta-Analysis of Mismatch Repair Polymorphisms within the Cogent Consortium for Colorectal Cancer Susceptibility', *PLoS ONE*, vol. 8, no. 9, e72091. <https://doi.org/10.1371/journal.pone.0072091>

### Digital Object Identifier (DOI):

[10.1371/journal.pone.0072091](https://doi.org/10.1371/journal.pone.0072091)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Publisher's PDF, also known as Version of record

### Published In:

PLoS ONE

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Meta-Analysis of Mismatch Repair Polymorphisms within the Cogent Consortium for Colorectal Cancer Susceptibility

Simone Picelli<sup>1,2,\*</sup>, Justo Lorenzo Bermejo<sup>3,4,9</sup>, Jenny Chang-Claude<sup>5</sup>, Michael Hoffmeister<sup>6</sup>, Ceres Fernández-Rozadilla<sup>7</sup>, Angel Carracedo<sup>7</sup>, Antoni Castells<sup>8</sup>, Sergi Castellví-Bel<sup>8</sup>, The EPICOLON Consortium<sup>1</sup>, Alessio Naccarati<sup>9</sup>, Barbara Pardini<sup>9</sup>, Ludmila Vodickova<sup>9,10</sup>, Heiko Müller<sup>11</sup>, Bente A. Talseth-Palmer<sup>12</sup>, Geoffrey Stibbard<sup>13</sup>, Paolo Peterlongo<sup>14,15</sup>, Carmela Nici<sup>14,15</sup>, Silvia Veneroni<sup>16</sup>, Li Li<sup>17</sup>, Graham Casey<sup>18</sup>, Albert Tenesa<sup>19</sup>, Susan M. Farrington<sup>19</sup>, Ian Tomlinson<sup>20</sup>, Victor Moreno<sup>21</sup>, Tom van Wezel<sup>22</sup>, Juul Wijnen<sup>23</sup>, Malcolm Dunlop<sup>19</sup>, Paolo Radice<sup>14,15</sup>, Rodney J. Scott<sup>12,24</sup>, Pavel Vodicka<sup>9,10</sup>, Clara Ruiz-Ponte<sup>7</sup>, Hermann Brenner<sup>11</sup>, Stephan Buch<sup>25</sup>, Henry Völzke<sup>26</sup>, Jochen Hampe<sup>25</sup>, Clemens Schafmayer<sup>27</sup>, Annika Lindblom<sup>1</sup>

**1** Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden, **2** Ludwig Institute for Cancer Research – Stockholm branch, Stockholm, Sweden, **3** Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg, Germany, **4** Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **5** Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **6** Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, **7** Galician Public Foundation of Genomic Medicine (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, Santiago de Compostela, University of Santiago de Compostela, Galicia, Spain, **8** Department of Gastroenterology, Hospital Clínic, The Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Catalonia, Spain, **9** Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic, **10** First Medical Faculty of the Charles University, Prague, Czech Republic, **11** Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, **12** School of Biomedical Science and Pharmacy, University of Newcastle, and the Hunter Medical Research Institute, Newcastle, Australia, **13** School of Science and IT, University of Newcastle, Newcastle, Australia, **14** Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, **15** Fondazione IFOM, Istituto FIRC di Oncologia Molecolare, Milan Italy, **16** Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy, **17** Department of Family Medicine, Case Center for Transdisciplinary Research on Energetics and Cancer, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio, United States of America, **18** University of Southern California, Norris Comprehensive Cancer Centre, Los Angeles, California, United States of America, **19** Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, University of Edinburgh and MRC Human Genetics Unit, Edinburgh, United Kingdom, **20** Oxford NIHR Comprehensive Biomedical Research Centre, Oxford, United Kingdom, **21** IDIBELL-Institut Català d'Oncologia (ICO), CIBER Epidemiología y Salud Pública (CIBERESP) and University of Barcelona, L'Hospitalet de Llobregat, Barcelona, Spain, **22** Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, **23** Department of Human Genetics and Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands, **24** Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW Australia, **25** Department of General Internal Medicine, University Hospital Schleswig-Holstein, Kiel, Germany, **26** Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, **27** Department of General and Thoracic Surgery, Christian-Albrechts-University, Kiel, Germany

## Abstract

In the last four years, Genome-Wide Association Studies (GWAS) have identified sixteen low-penetrance polymorphisms on fourteen different loci associated with colorectal cancer (CRC). Due to the low risks conferred by known common variants, most of the 35% broad-sense heritability estimated by twin studies remains unexplained. Recently our group performed a case-control study for eight Single Nucleotide Polymorphisms (SNPs) in 4 CRC genes. The present investigation is a follow-up of that study. We have genotyped six SNPs that showed a positive association and carried out a meta-analysis based on eight additional studies comprising in total more than 8000 cases and 6000 controls. The estimated recessive odds ratio for one of the SNPs, rs3219489 (MUTYH Q338H), decreased from 1.52 in the original Swedish study, to 1.18 in the Swedish replication, and to 1.08 in the initial meta-analysis. Since the corresponding summary probability value was 0.06, we decided to retrieve additional information for this polymorphism. The incorporation of six further studies resulted in around 13000 cases and 13000 controls. The newly updated OR was 1.03. The results from the present large, multicenter study illustrate the possibility of decreasing effect sizes with increasing samples sizes. Phenotypic heterogeneity, differential environmental exposures, and population specific linkage disequilibrium patterns may explain the observed difference of genetic effects between Sweden and the other investigated cohorts.

**Citation:** Picelli S, Lorenzo Bermejo J, Chang-Claude J, Hoffmeister M, Fernández-Rozadilla C, et al. (2013) Meta-Analysis of Mismatch Repair Polymorphisms within the Cogent Consortium for Colorectal Cancer Susceptibility. PLoS ONE 8(9): e72091. doi:10.1371/journal.pone.0072091

**Editor:** Nathan A. Ellis, University of Illinois at Chicago, United States of America

**Received:** January 10, 2013; **Accepted:** July 6, 2013; **Published:** September 6, 2013

**Copyright:** © 2013 Picelli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors are grateful to: the Spanish National Genotyping Center (CEGEN-ISCIII)-USC and the UPF nodes, as well as the Institutional Tumor Bank of Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. This study was supported by Instituto de Salud Carlos III, Spanish Ministry of Health, (grants FIS PI08-1635 and PS09-1037, VM); Cancer Research UK Programme Grant (C348/A12076) and a Centre Grant from the CORE Charity (MD); CZ:GA CR:GAP304/10/1286 and CZ:GA CR:GA310/07/1430 (PV); Hunter Medical Research Institute research grant and Priority Research Centre for Information Based Medicine (RS); German Research Council, German Federal Ministry of Research and Education (HB, HV); Baden Württemberg State Ministry of Research, Science and Arts (HB); Fondo de Investigación Sanitaria/FEDER (08/0024, 08/1276, PS09/02368, 11/00219, 11/00681); Instituto de Salud Carlos III (Acción Transversal de Cáncer), Xunta de Galicia (RH107/04 and 08CSA005208PR); Ministerio de Ciencia e Innovación (SAF2010-19273); Asociación Española contra el Cáncer (Fundación Científica y Junta de Barcelona); Fundació Olga Torres (CRP), and FP7 CHIBCHA Consortium (AC and SCB); Fondo de Investigación Sanitaria (CP 03-0070 to SCB and PS09/02368 to CFR). CIBERehd and CIBERER are funded by the Instituto de Salud Carlos III. JLB is partially supported by a grant of the Deutsche Forschungsgemeinschaft (DFG, SFB/TRR77, project Z2). LL is supported by a grant from National Cancer Institute R01 CA136726. SB, JH, WvS and CS were supported by the German Ministry for Education and Research through the German National Genome Research Network (NGFNplus) Colon Cancer Network (CCN). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: simone.picelli@lir.ki.se

† These authors contributed equally to this work.

‡ Membership of The EPICOLON Consortium is provided in the Acknowledgments.

## Introduction

In recent years low-risk common alleles have attracted increasing attention in the search for the “missing heritability” in colorectal cancer (CRC). It concerns the part of heritability that cannot be explained by mutations in already known high-risk genes but should, according to twin studies, account for about 35% [1]. Known high-penetrance germline mutations in CRC genes contribute for less than 6% of the observed cases [2]. Therefore, much of the remaining inherited variation in genetic susceptibility is probably due to multiple low-penetrance variants, both common and rare.

To date sixteen common variants have been identified through large multi-centre genome-wide association studies (GWAS) [3]. Taken together, however, they only explain a small proportion of familial CRC cases. Although the risk associated with each of these variants is modest, they contribute to the disease burden due to their high frequency in the population and the possibility of acting in concert with each other, which may increase the individual's risk of developing CRC [4].

Against this background, a few years ago we attempted to assess the role of eight SNPs in four already known CRC genes (*APC*, *MLH1*, *MSH6* and *MUTYH*) through a case-control association study in the Swedish population [5]. These 8 SNPs had been previously studied, but their pathogenicity was unknown and they were assumed to constitute polymorphisms. In our first study several positive associations were detected but, due to limited sample size (1785 cases and 1722 controls) [5], the results needed to be validated in a follow-up study.

The present study was an initiative of the COGENT consortium [4,6], where different groups offered to extend the genotyping to other non-Swedish cohorts for SNPs showing statistically significant associations in at least one analysis of the original study. This restricted the analysis to six out of the original eight SNPs.

## Materials and Methods

### Ethics statement

Collection of blood samples and clinical information from patients and controls was obtained with informed consent in accordance with the tenets of the Declaration of Helsinki. All participants gave written informed consent to take part in the study. The study was undertaken in accordance with the Swedish legislation of ethical permission (2003:460) and approved by the Stockholm Regional Research Ethical Committee (Dnr 2002:489).

### Mutation screening

Six SNPs in four different CRC genes were included in the analysis: rs459552:T>A (*APC* D1822V), rs1799977:A>G (*MLH1* I219V), rs1800932:A>G (*MSH6* P92P), rs1800935:T>C (*MSH6* D180D), rs3219484:G>A (*MUTYH* V22M) and rs3219489:G>C (*MUTYH* Q338H). *MUTYH* Q338H corresponds to Q324H in our first study [5]. The SNP nomenclature was modified to meet the Human Genome Variation Society's (HGVS) guidelines, which recommends the use of a reference sequence representing the largest theoretically known transcript. For *MUTYH* this corresponds to NM\_001128425.1 and NP\_001121897.1 for mRNA and protein, respectively [7,8,9].

### Subjects

Details regarding the number of cases and controls in all fourteen studies are summarized in Table S1. One SNP, rs459552 (*APC* D1822V), was genotyped in seven studies, for a total of 8654 cases and 7731 controls. Four SNPs, rs1799977 (*MLH1* I219V), rs1800932 (*MSH6* P92P), rs1800935 (*MSH6* D180D) and rs3219484 (*MUTYH* V22M) were genotyped in 8 studies for a total of 8308 cases and 7434 controls. The SNP with rs number 3219489 (*MUTYH* Q338H) was genotyped in 13 cohorts for a total of 12902 cases and 14602 controls.

For all the subjects genomic DNA was extracted from peripheral blood by standard procedures. Additional information regarding localization of the tumor, age at diagnosis, gender and ethnicity was retrieved whenever possible. Out of 5770 controls with ethnicity information, 5647 were of Caucasian origin, the rest being mostly African American.

### Genotyping

In studies 1, 5, 6, 7, 8, 9 and 10 SNPs were genotyped using the TaqMan SNP Genotyping Assay (Applied Biosystem, Foster City, CA). Genotyping in study 2 and 12 (controls only) was carried out by using the KASPar chemistry of the K-bioscience (Hoddesdon, Herts, UK) ([http://www.kbioscience.co.uk/reagents/KASP\\_manual.pdf](http://www.kbioscience.co.uk/reagents/KASP_manual.pdf)), which is a competitive allele-specific PCR SNP genotyping system that uses FRET quencher cassette oligos. Study 3 genotyped with the MassARRAY (Sequenom Inc., San Diego, USA) technology. Study 4 genotyped by means of fluorescent hybridization probe melting curves using the Light Cycler instrument (Roche). Study 11 genotyped using Illumina HumanHap 550 Bead Arrays. Study 12 was genotyped by Sanger sequencing (cases only). Studies 13 and 14 were genotyped using Illumina HumanHap300 and Illumina HumanHap240S.

## Statistical analysis

Deviations of observed genotype frequencies in controls from those expected under Hardy-Weinberg equilibrium were assessed by  $\chi^2$  tests. Risks of CRC associated with genotypes were compared by odds ratios (ORs) with corresponding confidence intervals (CIs) based on logistic regression. Study heterogeneity was summarized using a Mantel-Haenszel test but we assumed that the studies were random samples from a general population and used a random effect model to summarize OR estimates under dominant, recessive and additive penetrance models in the meta-analyses. Results were represented by forest plots as follows: confidence intervals for each individual study were indicated by horizontal lines, single ORs by squares and summary estimates by diamonds with horizontal limits at confidence limits and width inversely proportional to the standard error. Meta-analyses were performed using the package *rmeta* in the free software environment for statistical computing R.

## Results

The distribution of the genotypes in controls did not deviate from Hardy-Weinberg equilibrium in any study. Mantel-Haenszel tests identified study heterogeneity for rs1800932 (*MSH6* P92P) under recessive and additive penetrance, with p-values equal to 0.04 and 0.03, respectively (Table S2). This does not constitute a major issue since this SNP showed no differences between the genotype distributions of cases and controls either in single studies or in the global analysis. Study heterogeneity was not found for any other SNP. Genotyping results for the 6 SNPs based on studies 1–8 are presented in Table S2.

The only SNP that was marginally significant in the meta-analysis was rs3219489 (*MUTYH* Q338H), both under a recessive model (summary OR = 1.08, 95% CI 1.00 to 1.17;  $p = 0.05$ ) and assuming additive allelic effects (summary OR = 1.07, 95% CI 1.00 to 1.14;  $p = 0.06$ ). We ascribe the combined result mainly to the Swedish study, with individual ORs of 1.18 (95% CI = 1.01–1.38, recessive model) and 1.19 (95% CI = 1.05–1.35, additive model) (Table S2). The goodness of fit was slightly better for the recessive than for the additive model, and the recessive and additive models clearly outperformed the dominant model.

In an attempt to validate the findings under recessive inheritance, we set up collaborations with additional groups and requested to genotype rs3219489 in their cohorts. In the end, additional 4234 cases and 6800 controls were included, adding up to a total of 12232 cases and 13380 controls (Table S3).

We updated the meta-analysis once more considering all samples regardless of tumor localization as well as stratifying them for colon and rectal tumors. As shown in Table S4, data were available for 4573 colon and 1774 rectal cancer cases. Results from the updated meta-analyses are presented in Figure 1. The new summary OR for colorectal cancer was 1.03 (95% CI 0.97 to 1.10, probability value 0.25) (Figure 1A). The summary OR was practically identical after adjustment for age and gender OR = 1.03 (95% CI 0.93 to 1.13). Study heterogeneity was not noticed ( $P = 0.29$ , data not shown). The combined OR for colon cancer was 1.07 (95% CI 0.99 to 1.16, probability values 0.09 (OR = 1) and 0.37 (study homogeneity) (Figure 1B) and for rectal cancer was 1.06 (95% CI 0.94 to 1.19, probability values 0.37 (OR = 1) and 0.31 (study homogeneity)) (Figure 1C).

## Discussion

In the present investigation we performed a case-control association study for six out of eight previously investigated SNPs

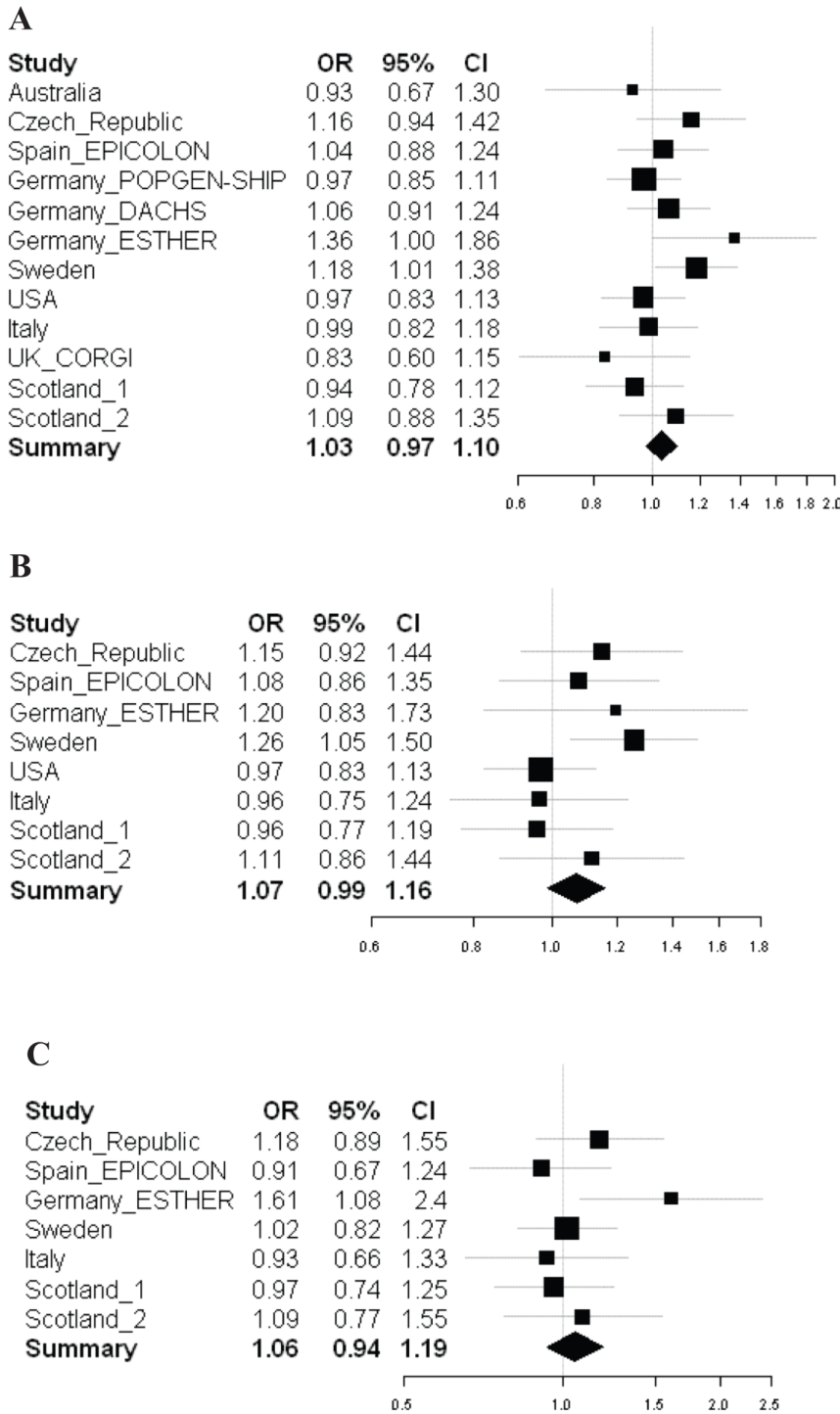
[5]. For five of them, rs459552 (*APC* D1822V), rs1799977 (*MLH1* I219V), rs1800932 (*MSH6* P92P), rs1800935 (*MSH6* D180D) and rs3219484 (*MUTYH* V22M) samples were retrieved from eight additional studies totaling 8308 cases and 7434 controls. For the sixth SNP, rs3219489 (*MUTYH* Q338H), which was selected based on promising results from two samples of Swedish origin (study 8 in the present manuscript and reference [5]), we set up an even larger replication dataset comprising 14 different studies with a total of 12232 cases and 13380 controls.

For all SNPs included in the analysis we were unable to confirm the associations with CRC risk found in the Swedish population. In particular, the recessive ORs of CRC for rs3219489 decreased from 1.52 in the original Swedish study to 1.18 in the Swedish replication cohort, to 1.08 (95% CI 1.00 to 1.17) in the first meta-analysis and to 1.03 (95% CI 0.97 to 1.10) in the updated meta-analysis (Table S2). The summary ORs in the extended meta-analyses were 1.07 (95% CI 0.99 to 1.16) for colon cancer and 1.06 (95% CI 0.94 to 1.19) for rectal cancer, in contrast with results based on Swedish samples. The updated meta-analysis had statistical power of 99% to detect a recessive OR of 1.52 and a power of 89% to detect a recessive OR of 1.18 (Type I error rate 5% and prevalence of CC genotypes among controls 5.6%). Biological plausibility was also existent. *MUTYH* Q338H is interesting because it represents a missense change in the *MUTYH* protein, which is involved in the base excision repair (BER) pathway. A common product of oxidative damage to 2'-deoxyguanosine is 7,8-dihydro-8-oxo-2'-deoxyguanosine (OG) [10,11]. In mammalian cells OG has been shown to be highly mutagenic and leading to an increased rate of G→T transversions, due to its miscoding properties that cause a mispairing with an adenine during DNA replication to form a stable OG:A mismatch [11,12]. The BER pathway plays an important role in repairing this type of DNA damage through the action of the mutY homolog *MUTYH*, in concert with *OGG1* and *MTH1* [11,13]. It is well established that biallelic mutations in *MUTYH* gene introduce G:C to T:A transversions also in the adenomatous polyposis coli (APC) gene, leading to genomic instability and abnormal and dysregulated cell proliferation in the colonic epithelium [14,15]. Patients with two mutations in the *MUTYH* gene develop the *MUTYH*-associated polyposis (MAP) syndrome [13].

To date, 85 different MAP-associated mutations have been found [16], scattered throughout the entire length of the protein, but only 3 (including Q338H) map within putative protein interaction domains as revealed by the recently solved crystal structure of h*MUTYH* [17]. It is tempting to speculate that Q338H might affect this protein-protein interaction, but additional experimental support is warranted.

The contrasting results on rs3219489 and its association with CRC risk in the Swedish versus other populations might suggest that the effect of this variant is specific for the Swedish population or not large enough in the other populations to be detected with the present sample size. For example, the statistical power of the updated meta-analysis was only 43% to detect a recessive OR of 1.10 (Type I error rate 5% and prevalence of CC genotypes among controls 5.6%). A closer look at the data actually shows that one of the German cohorts (ESTHER) gave results in agreement with our Swedish cohorts, with OR = 1.36 (95% CI 1.00 to 1.86) for colorectal cancer (Figure 1A) and OR = 1.61 (95% CI 1.08 to 2.40) for rectal cancer (Figure 1C). This is likely a spurious result due to the small size of that cohort (318 cases and 365 controls).

On the other hand, in agreement with Swedish results, rs3219489 has also been shown to be associated with CRC risk in three independent studies in the Japanese population [18,19,20] and among African-Americans (Yuan et al., 2nd InSiGHT



**Figure 1. Forest plots with observed odds ratios and 95% confidence intervals for rs3219489 (MUTYH Q338H) under a recessive penetrance model in colorectal cancer (A), colon cancer only (B) and rectal cancer only (C).**  
doi:10.1371/journal.pone.0072091.g001

meeting, Yokohama, Japan, unpublished) even though all these studies have a limited sample size and the results need further validation.

It is also possible that rs3219489 represents a risk-associated variant in the Swedish population in combination with environmental factors in the broad sense. For example, screening programs for CRC in Sweden could result in a diagnosis earlier

in life, thus inflating the ORs estimated in Sweden. Another alternative is that the polymorphism is in linkage disequilibrium with other unidentified causal variants. The marker and the causal variant could be located on the same risk haplotype in the Swedish population and on different haplotypes in other populations.

Independently of the unknown reason for replication failure, the results from the present study clearly illustrate the possibility of

decreasing effect sizes with increasing collections of individuals, a phenomenon well-known in the field of genetic epidemiology denominated the winner's curse [21]. It should be kept in mind that this outcome is rather expected in association studies, in particular those dealing with regionally heterogeneous complex diseases.

## Supporting Information

### Table S1 Number of cases and controls genotyped in the fourteen studies.

(DOC)

### Table S2 Genotype counts and allele frequencies for rs459552 (APC D1822V), rs1799977 (MLH I219V), rs1800932 (MSH6 P92P), rs1800935 (MSH6 D180D), rs3219484 (MUTYH V22M) and rs3219489 (MUTYH Q338H).

The estimated odds ratios with 95% confidence intervals for individual studies are also shown, together with combined ORs, 95% CIs and probability values for OR = 1 based on random effects model and probability values for study homogeneity under dominant, additive and recessive penetrance.

(DOC)

### Table S3 Genotype counts and allele frequencies for rs3219489 (MUTYH Q338H).

(DOC)

### Table S4 Genotype counts for colon and rectal cancer cases in studies with available information on tumor location.

(DOC)

## Acknowledgments

We thank all the patients that participated in this study. **Members of the EPICOLON Consortium (Gastrointestinal Oncology Group of the Spanish Gastroenterological Association):**

Hospital 12 de Octubre, Madrid: Juan Diego Morillas (local coordinator), Raquel Muñoz, Marisa Manzano, Francisco Colina, Jose Diaz, Carolina Ibarrola, Guadalupe López, Alberto Ibáñez; Hospital Clínic, Barcelona: Antoni Castells (local coordinator), Virginia Piñol, Sergi Castellví-Bel, Francesc Balaguer, Victoria Gonzalo, Teresa Ocaña, María Dolores Giráldez, María Pellisé, Anna Serradesanferm, Leticia Moreira, Miriam Cuatrecasas, Josep M. Piqué; Hospital Clínic Universitario, Zaragoza: Ángel Lanás (local coordinator), Javier Alcedo, Javier Ortega; Hospital Cristal-Piñor, Complejo Hospitalario de Ourense: Joaquín Sánchez (local coordinator), M<sup>a</sup> Soledad Díez, Mercedes Salgado, Eloy Sábchez, Mariano Vega; Parc de Salut Mar, Barcelona: Montserrat Andreu (local coordinator), Anna Abuli, Xavier Bessa, Mar Iglesias, Agustín Seoane, Felipe Bory, Gemma Navarro, Beatriz Bellosillo; Josep M<sup>a</sup> Dedeu, Cristina Álvarez, Marc Puigvehí; Hospital San Eloy, Baracaldo and Hospital Donostia, CIBERehd, University of Basque Country, San Sebastián: Luis Bujanda (local coordinator) Ángel Cosme, Inés Gil, Mikel

Larzabal, Carlos Placer, María del Mar Ramírez, Elisabeth Hijona, Jose M. Enríquez-Navascués, Jose L. Elosegui; Hospital General Universitario de Alicante: Artemio Payá (EPICOLON I local coordinator), Rodrigo Jover (EPICOLON II local coordinator), Cristina Alenda, Laura Sempere, Nuria Acame, Estefanía Rojas, Lucía Pérez-Carbonell; Hospital General de Granollers: Joaquim Rigau (local coordinator), Ángel Serrano, Anna Giménez; Hospital General de Vic: Joan Saló (local coordinator), Eduard Batiste-Alentorn, Josefina Autonell, Ramon Barniol; Hospital General Universitario de Guadalajara and Fundación para la Formación e Investigación Sanitarias Murcia: Ana María García (local coordinator), Fernando Carballo, Antonio Bienvenido, Eduardo Sanz, Fernando González, Jaime Sánchez, Akiko Ono; Hospital General Universitario de Valencia: Mercedes Latorre (local coordinator), Enrique Medina, Jaime Cuquerella, Pilar Canelles, Miguel Martorell, José Ángel García, Francisco Quiles, Elisa Orti; CHUVI-Hospital Meixoeiro, Vigo: EPICOLON I: Juan Clofent (local coordinator), Jaime Seoane, Antoni Tardío, Eugenia Sanchez. EPICOLON II M<sup>a</sup> Luisa de Castro (local coordinator), Antoni Tardío, Juan Clofent, Vicent Hernández; Hospital Universitari Germans Trias i Pujol, Badalona and Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, IL, USA: Xavier Llor (local coordinator), Rosa M. Xicola, Marta Piñol, Mercè Rosinach, Anna Roca, Elisenda Pons, José M. Hernández, Miquel A. Gassull; Hospital Universitari Mútua de Terrassa: Fernando Fernández-Bañares (local coordinator), Josep M. Viver, Antonio Salas, Jorge Espinós, Montserrat Forné, María Esteve; Hospital Universitari Arnau de Vilanova, Lleida: Josep M. Reñé (local coordinator), Carmen Piñol, Juan Buenestado, Joan Viñas; Hospital Universitario de Canarias: Enrique Quintero (local coordinator), David Nicolás, Adolfo Parra, Antonio Martín; Hospital Universitario La Fe, Valencia: Lidia Argüello (local coordinator), Vicente Pons, Virginia Pertejo, Teresa Sala; Hospital Sant Pau, Barcelona: Dolores Gonzalez (local coordinator) Eva Roman, Teresa Ramon, María Poca, M<sup>a</sup> Mar Concepción, Marta Martín, Lourdes Pétriz; Hospital Xeral Cies, Vigo: Daniel Martínez (local coordinator); Fundación Pública Galega de Medicina Xenómica (FPGMX), CIBERER, Genomic Medicine Group-University of Santiago de Compostela, Santiago de Compostela, Galicia, Spain: Ángel Carracedo (local coordinator), Clara Ruiz-Ponte, Ceres Fernández-Rozadilla, M<sup>a</sup> Magdalena Castro; Hospital Universitario Central de Asturias: Sabino Riestra (local coordinator), Luis Rodrigo; Hospital de Galdácano, Vizcaya: Javier Fernández (local coordinator), Jose Luis Cabriada; Fundación Hospital de Calahorra (La Rioja) La Rioja: Luis Carreño (local coordinator), Susana Oquiñena, Federico Bolado; Hospital Royo Villanova, Zaragoza: Elena Peña (local coordinator), José Manuel Blas, Gloria Ceña, Juan José Sebastián; Hospital Universitario Reina Sofía, Córdoba: Antonio Naranjo (local coordinator).

## Author Contributions

Conceived and designed the experiments: AL. Performed the experiments: SP JCC MH CFR A. Carracedo A. Castells SCB AN BP LV HM BTP GS PP CN SV LL GC AT SMF IT VM TVW JW MD PR RJS PV CRP HB SB HV JH CS. Analyzed the data: SP JLB. Contributed reagents/materials/analysis tools: AL JLB JCC MH CFR A. Carracedo A. Castells SCB AN BP LV HM BTP GS PP CN SV LL GC AT SMF IT VM TVW JW MD PR RJS PV CRP HB SB HV JH CS EPICOLON Consortium. Wrote the paper: SP JLB AL.

## References

- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, et al. (2000) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343: 78–85.
- 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: 356–361.
- Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet* 42: 973–977.
- Tomlinson IP, Dunlop M, Campbell H, Zanke B, Gallinger S, et al. (2010) COGENT (COLOrectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer. *Br J Cancer* 102: 447–454.
- Picelli S, Zajac P, Zhou XL, Edler D, Lenander C, et al. (2010) Common variants in human CRC genes as low-risk alleles. *Eur J Cancer* 46: 1041–1048.
- Houlston RS (2012) COGENT (COLOrectal cancer GENeTics) revisited. *Mutagenesis* 27: 143–151.
- den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 15: 7–12.
- den Dunnen JT, Paalman MH (2003) Standardizing mutation nomenclature: why bother? *Hum Mutat* 22: 181–182.
- Out AA, Tops CM, Nielsen M, Weiss MM, van Minderhout IJ, et al. (2010) Leiden Open Variation Database of the MUTYH gene. *Hum Mutat* 31: 1205–1215.
- Klaunig JE, Kamendulis LM (2004) The role of oxidative stress in carcinogenesis. *Annu Rev Pharmacol Toxicol* 44: 239–267.
- David SS, O'Shea VL, Kundu S (2007) Base-excision repair of oxidative DNA damage. *Nature* 447: 941–950.

12. Neeley WL, Essigmann JM (2006) Mechanisms of formation, genotoxicity, and mutation of guanine oxidation products. *Chem Res Toxicol* 19: 491–505.
13. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, et al. (2002) Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 30: 227–232.
14. Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, et al. (2003) Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 362: 39–41.
15. Sieber OM, Lipton L, Crabtree M, Heinimann K, Fidalgo P, et al. (2003) Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. *N Engl J Med* 348: 791–799.
16. Cheadle JP, Sampson JR (2007) MUTYH-associated polyposis—from defect in base excision repair to clinical genetic testing. *DNA Repair (Amst)* 6: 274–279.
17. Luncsford PJ, Chang DY, Shi G, Bernstein J, Madabushi A, et al. (2010) A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9-Rad1-Hus1 checkpoint complex interactions. *J Mol Biol* 403: 351–370.
18. Tao H, Shimura K, Suzuki M, Kono S, Mibu R, et al. (2008) Association between genetic polymorphisms of the base excision repair gene MUTYH and increased colorectal cancer risk in a Japanese population. *Cancer Sci* 99: 355–360.
19. Kasahara M, Osawa K, Yoshida K, Miyaishi A, Osawa Y, et al. (2008) Association of MUTYH Gln324His and APEX1 Asp148Glu with colorectal cancer and smoking in a Japanese population. *J Exp Clin Cancer Res* 27: 49.
20. Yanaru-Fujisawa R, Matsumoto T, Ushijima Y, Esaki M, Hirahashi M, et al. (2008) Genomic and functional analyses of MUTYH in Japanese patients with adenomatous polyposis. *Clin Genet* 73: 545–553.
21. Ioannidis JP, Thomas G, Daly MJ (2009) Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 10: 318–329.