

## ANSWERING REVIEWERS



December 19th, 2013

Dear Editor,

Please find enclosed the edited manuscript in Word format (file name: 6073-New genes emerging for colorectal cancer predisposition\_edited2.doc).

**Title:** New genes emerging for colorectal cancer predisposition

**Author:** Clara Esteban-Jurado, Pilar Garre, Maria Vila, Juan José Lozano, Anna Pristoupilova, Sergi Beltrán, Anna Abulí, Jenifer Muñoz, Francesc Balaguer, Teresa Ocaña, Antoni Castells, Josep M Piqué, Angel Carracedo, Clara Ruiz-Ponte, Xavier Bessa, Montserrat Andreu, Luis Bujanda, Trinidad Caldés, Sergi Castellví-Bel

**Name of Journal:** *World Journal of Gastroenterology*

**ESPS Manuscript NO:** 6073

Please find attached a revised version our manuscript to be considered for publication in the *World Journal of Gastroenterology* as a Review Article. We are grateful for the opportunity to resubmit our article.

The manuscript has been modified to reflect the helpful and insightful comments provided by both reviewers. We hope our manuscript will be now in accordance with the editor and reviewers' comments and suitable for publication in its present form. We have addressed all comments made by editor and reviewers and we are answering them giving a point-by-point response to their concerns in the following pages. We have also highlighting all changes made in the manuscript to make it easier to follow them (yellow for reviewers 1 and 2, green for reviewer 3).

This manuscript has been seen and approved by all authors, who have taken due care to ensure the integrity of the work. The authors warrant full disclosure of any financial and personal relationships with other people or organizations that could inappropriately influence their work.

Thank you again for publishing our manuscript in the *World Journal of Gastroenterology*.

Sincerely yours,

A handwritten signature in blue ink, appearing to read 'Sergi Castellví-Bel'.

Sergi Castellví-Bel, Ph.D.

Genetic Susceptibility to Colorectal Cancer Group

Gastrointestinal & Pancreatic Oncology Team

IDIBAPS / CIBERehd / Hospital Clínic

Centre Esther Koplowitz (CEK)

Rosselló 153 planta 4

08036 Barcelona; Spain

34-93-2275400 ext.4183/2915

[sbel@clinic.ub.es](mailto:sbel@clinic.ub.es)

Chair for COST Action BM1206

<http://www.eucolongene.eu/>

## RESPONSE TO EDITOR

1. We are aware of your policy regarding the revision of invited review articles that do not reach an academic quality level of grade B or above after peer review. However, we would like to point out that one of the reviewers already rated our review as grade B whereas the other one gave us a C. With all improvements made, we are confident we can achieve now a better grade.

2. We are also aware of your policy regarding language evaluation of articles, stating that for manuscripts submitted by non-native speakers of English, the author(s) must seek to make use of a copyediting service provided by professional English language editing companies. However, we would like to point out that one of the reviewers already rated our review as grade A in this area whereas the other one gave us a B. With all corrections made, we are confident we can now achieve an A grade from both. Therefore, we thank you for your suggestion but we believe that it is not necessary to use this service. As corresponding author and final responsible of this review, I would like to point out that I have authorship in 99 Pub-Med indexed articles written in English and I have never needed or been required by the publishers to make use of such service. Please, take also into account that nowadays research funding is sparse in Spain and we do not spend any budget on these matters. We actually were able to accept to contribute with this review since you offered to publish it free of charge.

3. Finally, we have addressed the editor's comment regarding cities postcodes.

---

## RESPONSE TO REVIEWER 1 (131028)

### General Comments

1. The authors present a review paper on new genes emerging for colorectal cancer (CRC) predisposition, with focus on new sequencing technologies. This is an important topic of its field. One main finding is related to germline mutations in the POLE and POLD1 genes responsible for a new form of predisposition called polymerase proofreading-associated polyposis. It is to be noted that this has partly been highlighted in literature elsewhere (Seshagiri 2013).

*As suggested, this reference has been added on page 16 and the list of References:*

*Seshagiri S. The burden of faulty proofreading in colon cancer. Nat Genet 2013; 45: 1261 [PMID: 24071854 DOI: 10.1038/ng1013-1261b].*

2. The authors spend time on reviewing how next generation sequencing (NGS) technologies may be used to reveal germline variations predisposing CRC. There is a separate paragraph on how individuals should be selected for this analysis, although this seems to be related to hereditary CRC cases only. Of the NGS studies that are presented in the manuscript, the study designs are different and as such patients have been selected differently. For instance the study by Palles et al. included patients with a positive history of CRC, whereas the study by Smith et al. included unrelated patients with sporadic CRC. When presenting the CDCV hypothesis, this can also be seen as an effort to explain genetic susceptibility of sporadic CRC.

Thus, when reading the manuscript it is not clear to the reader if the focus is on hereditary, familial and/or sporadic CRC. This review paper should make this clear, and further discuss implications of the different study designs related to the inclusion of patients with or without a family history of CRC.

*We would like to point out that in this article we intended to review genetic predisposition to colorectal cancer. In that sense, the common disease-common variant is commented when a polygenic model of inheritance is considered for colorectal cancer, which could be involved both in sporadic and familial forms of this disease.*

*On the other hand, the objective of the “Overview of new sequencing technologies” section of the review was to offer a glimpse of this technical breakthrough and how it can be applied in the identification of genes involved in predisposition to human disease in general and its application to colorectal cancer only is avoided. Therefore, we believe the objective of this review was not to look at this subject thoroughly. However, as suggested by both reviewers, we have clarified this section when explaining the different study designs for inclusion of patients, either familiar or sporadic, as it is shown below:*

*“For diseases with genetic heterogeneity as human cancers, different strategies can be used including the selection of families with strong disease aggregation or sequencing sporadic cases with early onset for the disease. Both situations are suggestive of the involvement of a germline predisposition. When focusing in families with several affected members, sequencing can be performed in several cases in each family and only those shared variants will be taken into account. On the other hand, if sporadic early-onset cases are chosen, genes with variants in different individuals can be selected.”*

3. In conclusion, the manuscript should be re-structured and formatted to make it more concise, to the point and easy to read. There are also several typographic errors in the manuscript, some of which are mentioned below.

*As suggested by both reviewers, we have made amendments especially in the “Overview of new sequencing technologies” section to make it clearer to the reader. Also, typographic errors have been corrected.*

### **Some Specific Comments**

#### **Introduction**

- Page 5, 1st paragraph: I suggest estimates are presented with less significant figures, for instance 473,258 could be presented as approximately 473,200.

*As suggested, we now present estimates with less significant figures:*

*“For 2015, approximately 473,200 new cases are predicted and 233,900 individuals will die from this disease in Europe.”*

- Page 5, 1st paragraph: Incidence rates are compared across the world without a comment on the quality variations in cancer registries.

*As suggested by the reviewer, we now state:*

*“The incidence of CRC varies widely between countries, depending on their degree of development and also on the quality of their cancer registries.”*

- Page 5, 2nd paragraph: There is a repetition of the prevalence from the 1st paragraph: “The lifetime risk of CRC in the general population is about 5-6 % in Western countries..” vs “Approximately 5 % of the population develops..”. This figure should be stated consistently.

*Colorectal cancer prevalence is now stated consistently as 5%.*

#### **Hereditary CRC**

- Page 8, 3rd paragraph: "Most patients have a family history of colorectal polyps and cancer, but de novo APC mutations are responsible for approximately 25 % of cases". This lacks a reference in the manuscript.

*As suggested, this reference has been added on page 9 corresponding to a reference already included (Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology 2010; 138: 2044-58 [PMID: 20420945 DOI:10.1053/j.gastro.2010.01.054]).*

- Clinical literature such as UpToDate (accessed October 2013) often speaks of familial adenomatous polyposis (FAP) and MYH-associated polyposis (MAP) as two separate entities caused by different germline mutations, although their clinical spectrum might overlap. I therefore suggest that there is a separate paragraph on MAP. Furthermore I suggest Table 1 to be altered accordingly so that the syndromes FAP and MAP are separated.

*We would like to point out that in this article we intended to review genetic predisposition to colorectal cancer. However, the main objective of this review was to focus on new genes involved recently in predisposition to this disease. Therefore, we have added a sentence explaining that for classical hereditary colorectal cancer syndromes, we only revise concisely the more common forms, being hereditary nonpolyposis colorectal cancer and familial adenomatous polyposis, as stated below:*

*"The most frequent forms are hereditary nonpolyposis colorectal cancer and familial polyposis syndrome, which are further described below."*

Approaches to identify genetic variants for CRC risk

- Page 9, 3rd paragraph: Typographic error "common disease-common variant" .

*It has been corrected.*

Overview of new sequencing technologies

- Is this section the focus of the paper? It should be re-organized somewhat because it both discusses the selection of patients (pre-analytical), sequencing methods and filtering processes (analytical) and validation of candidate variants (post-analytical) in an unorganized fashion.

- I suggest creating a figure illustrating the step-wise process including important preanalytical, analytical and post-analytical steps.

*It is not the main focus of this review. As previously commented, the objective of the "Overview of new sequencing technologies" section was to offer a glimpse of this technical breakthrough and how it can be applied in the identification of genes involved in predisposition to human disease. Therefore, we believe that it is already easier to read for a non-expert in its present form with the added corrections and it is not necessary to add any more subsections or figures.*

- In general pages 10 – 12 would benefit from more references.

- Page 10, 2nd paragraph: "Due to recent technology and variant calling algorithm improvements, NGS is probably nowadays more accurate than Sanger sequencing". This lacks a reference in the manuscript.

*A new reference has been added as suggested:*

Ullahannan D, Kovac MB, Mulholland PJ, Cazier JB, Tomlinson I. Technical and implementation issues in using next-generation sequencing of cancers in clinical practice. *Br J Cancer* 2013; 109: 827-35 [PMID: 23887607 DOI: 10.1038/bjc.2013.416].

- Page 11, 3rd paragraph: "...which allows sequencing a larger number of samples with better accuracy". How do you define the term accuracy in this context?

*As suggested by the reviewer, a clarification for accuracy in this context has been added:*

*"One advantage of WES is that is about much cheaper than WGS, which allows sequencing a larger number of samples with better accuracy or coverage. The term coverage corresponds to the read depth or depth and it is the average number of times that a nucleotide has been sequenced in a different sequencing read."*

- Page 11, 4th paragraph: This paragraph is discussing selection of patients (preanalytical) and is obviously a critical step in the design of such studies. This part could be presented first in this section. In addition it only describes the inclusion of patients with a positive CRC history. It also lacks considerations regarding the number of patients needed in order to find given genetic variations (i.e. the power).

*As suggested by both reviewers, we have clarified this section when explaining the different study designs for inclusion of patients, either familiar or sporadic. Also, regarding the reviewer's comment about power, we have incorporated a comment about the possibility to obtain good results with next generation sequencing when using carefully selected patients in contrast to genetic association studies where number of cases and controls needs to be much higher, as stated below:*

*"Also, it should be noted that is possible to obtain good results with NGS when using carefully selected patients in contrast to GWAS, where number of cases and controls that are compared needs to be much higher in order to obtain statistically significant findings."*

- Page 11, sub-section "Data filtering and prioritization in NGS": Why selecting a coverage threshold of 10 x - is there any literature supporting this notion?

*As suggested by the reviewer, we now state that coverage threshold to avoid sequencing artifacts is "typically 5-10x", increasing its range to reflect many of the available published studies.*

- Page 13, 8th paragraph: Sanger sequencing is only one of several possibilities for low throughput sequencing validation.

*We agree with the reviewer and we have added a sentence as stated below:*

*"Sequencing validation by Sanger sequencing or any other PCR technology designed to detect a specific nucleotide change is necessary after NGS to confirm the prioritized variants and exclude sequencing artifacts."*

New genes identified for CRC genetic predisposition

- This section refers to the results of both low-throughput sequencing and highthroughput sequencing studies, and is initially not clear to the reader.

*We have added a sentence at the begining explaining that both low-throughput sequencing and NGS was used in the described studies, as stated below:*

*"New sequencing technologies made available recently including exome- and whole-genome sequencing have permitted to add a new approach to facilitate the identification of new genes responsible for human disease predisposition. Indeed, some seminal efforts have been already completed very recently for CRC. However, before these high-throughput technologies have yielded results in CRC families, some previous low-throughput sequencing studies reported directed screening of some plausible gene candidates for various reasons."*

#### Figures/Tables

- Figure 1: Incorrect reference name in figure legend:

*"(data adapted from Ferlay et al. [1].)"*

*It has been corrected.*

- Table 1: Lacks reference in table legend.

*It has been added.*

- Table 2: Lacks reference in table legend.

*They have been added.*

#### References

Seshagiri S. Nat Genet. 2013 Feb;45(2):121-2. doi: 10.1038/ng.2540.

*It has been added.*

---

## RESPONSE TO REVIEWER 2 (02445033)

### **Comments to authors:**

The manuscript deals with the description of new genes that may increase the risk of developing a CRC. It is well written and references are much updated. The core of the manuscript seems to be a description of new sequencing technologies and their applicability on the search of new CRC predisposition genes. However, the general structure of the manuscript seems a little confusing to the reader. It begins focusing on hereditary syndromes and then jumps to the new sequencing technologies, when following the text these technologies seem to be applied mainly to familial CRCs. The authors comment on the election of individuals for sequencing on page 11, addressing the importance of choosing families with stronger familial aggregation, but the published reports on this topic include both sporadic and familial cases. Perhaps a more specific description of selection of individuals for sequencing and its influence on the results of the studies should clarify the practical relevance of these findings.

*As previously mentioned, the objective of the "Overview of new sequencing technologies" section of the review was to offer a glimpse of this technical breakthrough and how it can be applied in the identification of genes involved in predisposition to human disease in general and its application to colorectal cancer only is avoided. However, as suggested by both reviewers, we have clarified this section when explaining the different study designs for inclusion of patients, either familiar or sporadic, as it was previously shown.*

As a general review on new candidate genes, and from a non-genetist reader point of view, I would like some comment on what should be the next step to elucidate the pathogenic role of these genes. We are overwhelmed with hundreds of putative genes but most of them may not be real driver genes. How should we deal with all this information from a translational perspective?

*As suggested by the reviewer, we have explained more precisely the next steps to elucidate the pathogenicity of candidate genes/variants, as it is shown below:*

*"Sequencing validation by Sanger sequencing or any other PCR technology designed to detect a specific nucleotide change is necessary after NGS to confirm the prioritized variants and exclude sequencing artifacts. Also, segregation analysis in families permits to check if a candidate variant segregates correctly with the disease. Therefore, affected members need to be carriers and non-affected individuals old enough to be expressing the disease should be non-carriers in order to find correct segregation of the candidate variant with the studied disease. Additionally in the case of hereditary cancer, when heterozygous candidate variants with correct segregation are identified, it is necessary to confirm if there is loss of the second allele in the tumor DNA in order to establish the candidate gene as a tumor suppressor gene. Case-control screening studies can also be performed in order to identify additional carriers of the candidate variants in ample disease cohorts and further demonstrate its absence in controls. Finally, functional assessment of the candidate variant and affected gene will be also necessary to further confirm the negative effect of the variant in the protein and prove its involvement in disease development by in vitro studies and animal models."*

Finally some typographic mistakes: - Page 8, cutaneous, sebaceous - Page 9, common disease-common variant.

*They have been corrected.*

---

## RESPONSE TO REVIEWER 3 (02854680)

### **Comments to authors:**

This was a good review article for the general clinician with no previous in depth knowledge of genetics. The review gives an important overview on the genetic methods and pros and cons of next gen, whole genome versus exome sequencing. Although many of the findings discussed in this article are not particularly new or novel, they provide a sound background for the general reader. Perhaps the article should have a larger focus on NEW findings in CRC genetics.

### **Major suggestions:**

Final section Expand on the functional role of POLE and POLD-1 i.e. polymerase proofreading genes.

*As suggested by the reviewer, we have expanded this part, as it is shown below:*

*"POLE and POLD1 encode the catalytic and proofreading activities of the leading-strand DNA polymerase  $\epsilon$  and the lagging-strand polymerase  $\delta$ . The proofreading capacity of the exonuclease domain is essential for the maintenance of replication fidelity and may act not only on newly misincorporated bases but also on mismatches produced by non-proofreading polymerases."*

Has any other functional in-or ex vivo work been done on these genes?

*As suggested by the reviewer, we have expanded this part, as it is shown below:*

*“Functional assessment supported the importance of these mutations in POLE and POLD1. Mutagenesis studies of Pol $\delta$  and Pol3 in yeast showed that the mutation of the equivalent residue produces a mutator phenotype and loss of the proofreading activity of the protein<sup>[48,50,51]</sup>. Also, mice expressing proofreading-impaired Pole and Pold1 in a homozygous state developed spontaneous intestinal adenocarcinomas or a spectrum of cancers<sup>[52]</sup>.”*

Table1- could the functions of the genes included be another column of this table i.e. tumor suppressor gene, mismatch repair gene, unknown etc...

*As suggested by the reviewer, we have expanded this table, as it is shown below:*

*Table 1. Hereditary colorectal cancer genes.*

	Gene	Chromosome	Mendelian pattern	Function
Familial adenomatous polyposis	APC	5q	AD	regulation of canonical Wnt signaling pathway
	MUTYH	1p	AR	base-excision repair
Hereditary non-polyposis CRC (Lynch syndrome)	MLH1	3p	AD	mismatch repair
	MSH2	2p	AD	mismatch repair
	MSH6	2p	AD	mismatch repair
	PMS2	7p	AD	mismatch repair
Peutz-Jeghers	LKB1	19p	AD	regulation of Wnt signaling pathway
Juvenile polyposis	SMAD4	18q	AD	TGFBR signaling pathway
	BMPR1A	10q	AD	TGFBR signaling pathway
Cowden's disease	PTEN	10q	AD	negative regulation of PI3K signaling
	KLLN	10q	AD	apoptotic process

CRC, colorectal cancer; AD, autosomal dominant; AR, autosomal recessive; TGFBR, transforming growth factor beta receptor; PI3K, phosphatidylinositol 3-kinase

Reference: Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. *Gastroenterology* 2010; 138: 2044-58.

## Minor suggestions:

Sentence 4 of the abstract (also appears in core tip) starting “Therefore, there is...” needs to be revised as it makes little sense.

*As suggested by the reviewer, we have revised this sentence both in the abstract and the core tip, as it is shown below:*

*“Excluding hereditary forms, there is an important fraction of CRC cases that present familial aggregation for the disease with an unknown germline genetic cause.”*

Sentence 3 of the introduction change “For 2015” to “In 2015”

*It has been corrected as suggested.*

More references are required for the introduction as there are many unreferenced comments/evidence/statistics.

*Six more references have been added to this article as suggested.*

In the paragraph regarding Lynch Syndrome- the epigenetic aspects are interesting and should be expanded upon.

*As suggested by the reviewer, we have expanded this part, as it is shown below:*

*“Recently, germline deletions of the 3’ region of EPCAM gene were found in a subset of families with Lynch syndrome. This deletion leads to promoter hypermethylation of MSH2, located upstream of the deleted gene<sup>[16]</sup>. The MMR system is necessary to maintaining genomic fidelity by correcting single-base mismatches and insertion-deletion loops during DNA replication. As a consequence, Lynch syndrome tumors accumulate errors in short repetitive sequences, a phenomenon called microsatellite instability (MSI), which is considered a landmark for this disease. It is noteworthy to mention that in sporadic MSI CRC cancers, loss of expression of MLH1 due to hypermethylation of its promoter is a frequent event, and it is linked with the somatic mutation V600E in the BRAF gene<sup>[17]</sup>.”*

Please expand on the Amsterdam criteria (perhaps a figure or a table if necessary)

*As suggested by the reviewer, we have expanded this part into a table, as it is shown below:*

**Table 2.** The Amsterdam criteria in Lynch syndrome.

<b>Amsterdam Criteria I</b>
-----------------------------

*At least three relatives with CRC; all of the following must be met:*

- 1. One affected individual is a first degree relative of the other two*
- 2. At least two successive generations affected*
- 3. At least one CRC diagnosed before the age of 50 years*
- 4. Familial adenomatous polyposis has been excluded*

## **Amsterdam Criteria II**

*At least three relatives with colorectal, endometrial, small bowel, ureter, or renal pelvis cancer; all of the following must be met:*

- 1. One affected individual is a first degree relative of the other two*
- 2. At least two successive generations affected*
- 3. At least one tumor diagnosed before the age of 50 years*
- 4. Familial adenomatous polyposis has been excluded*

First sentence page 10-change nonsyndromic to non-syndromic.

*It has been corrected.*

Sentence one page 11 “de novo” instead of “the novo”

*It has been corrected.*

Final sentence of paragraph 2 of page 11, “different systematic error associated, as conventional Sanger sequencing, which increases the costs and time of the analysis.” Insert such after comma.

*It has been corrected.*

Page 13 expand abbreviation SNV

*It has been expanded.*

When discussing FAP- what about peri-ampullary tumours?

*As suggested by the reviewer, we have expanded this part, as it is shown below:*

*“Adenomatous polyps are also found in the stomach and duodenum, especially the periampullary area and can develop into adenocarcinomas. After colectomy, periampullary carcinoma is the most common malignancy, occurring in approximately 5-6% of the patients<sup>[19]</sup>.”*