Bukti Submit Artikel ke System Pharmacological Resarch

https://www.editorialmanager.com/yphrs/default.aspx

editorialmanager.com/yphrs/default.aspx

Pharmacological RESEARCH

Editorial Manager

HOME • LOGOUT • HELP • REGISTER • UPDATE MY INFORMATION • JOURNAL OVERVIEW
MAIN MENU • CONTACT US • SUBMIT A MANUSCRIPT • INSTRUCTIONS FOR AUTHORS • PRIVACY

Role: Author Username: lalu

Submissions with an Editorial Office Decision for Author lalu Muhammad Irham

Page: 1 of 1 (3 total completed submissions) Display 10 results per page.

Action	Manuscript Number	Title	Authorship	Initial Date Submitted	Status Date	Current Status	Date Final Disposition Set	Final Disposition
Action Links	YPHRS-D-20-01156	Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer	Other Author	Jun 06, 2020	Sep 10, 2020	Completed - Accept	Sep 10, 2020	Accept

Page: 1 of 1 (3 total completed submissions) Display 10 results per page.

<< Author Main Menu

You should use the free Adobe Reader 10 or later for best PDF Viewing results.

Get ADOBE READER

ELSEVIER

Log In Register Help

Track Your Accepted Article

The easiest way to check the publication status of your accepted article

Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer

Article reference	YPHRS_105203
Journal	Pharmacological Research
Corresponding author	Wei-Chiao Chang
First author	lalu Muhammad Irham
Received at Editorial Office	6 Jun 2020
Article revised	27 Aug 2020
Article accepted for publication	9 Sep 2020
DOI	10.1016/j.phrs.2020.105203

ISSN 1043-6618

Last update: 12 Oct 2020 Share via email

Production events

12 Oct 2020	The Share Link has been sent to you
12 Oct 2020	Final version of your article published online
4 Oct 2020	Your proof corrections have been returned to Elsevier
4 Oct 2020	Proofs available for checking
21 Sep 2020	Rights & Access form completed by you
21 Sep 2020	Offprint order form completed by you
17 Sep 2020	Article published online
13 Sep 2020	Rights & Access form sent to you for completion
13 Sep 2020	Offprint order letter sent to you for completion
10 Sep 2020	Received for production

Full bibliographic details
Pharmacological Research 161C (2020) 105203

Final article available online
12 Oct 2020

View your final article

Cited by in Scopus: 0

Get your Stats on Mendeley

Track another article

Now also available

Feedback

Feedback

Response to Reviewer Comment:

Reviewer 1: Manuscript Title: Integration of Genetic Variants and Gene Network for Drug Repositioning in Colorectal Cancer (Lalu Muhammad Irham1,2,3, Henry Sung-Ching Wong1,2, Wan-Hsuan Chou1,2, Wirawan Adikusuma1,2,4, Eko Mugiyanto1,2,5, Wan-Chen Huang6*, Wei-Chiao Chang1,2,7,8,9*)

General Comments:

The manuscript entitled "Integration of Genetic Variants and Gene Network for Drug Repositioning in Colorectal Cancer" is comprehensive study of the drug repurposing in colorectal cancer generally. Title of article is appropriate. Abstract is good arranged. Introduction is good organized with sufficient literature review. Methods are evidence based. Results show findings clearly. Conclusion is logical and appropriate. References are related to the issue. I think the figures, tables and highlights are important and help to better understanding of the subject although I have some corrections to improve the article value that I describe in "Minor & Major comments".

Answer: We thank the reviewer's comments

Minor Comments:

Q1: Reviewer #1: Please change "Repositioning" to "Repurposing" in "title" of article.
A1: We thank the reviewer's comments. **Yes, we have revised the word: "Repositioning" to "Repurposing" in the title of the manuscript according to the reviewer's suggestions. As shown in the title "Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer". [Page 1, lines 3-4]**

Q2: Reviewer #1: Please add related reference(s) after "drugs." in line 103.
A2: Thanks for the reviewer's suggestions. **We have revised the manuscript by adding two references [16 and 17]. These two references are shown as below, and we added them accordingly in the manuscript. [Page 4, line 103]**

[16] E.M. Stoffel, E. Koeppe, J. Everett, *et al.*, Germline Genetic Features of Young Individuals With Colorectal Cancer, Gastroenterology 154(4) (2018) 897-905.e1.
[17] S.A. Smith, T. French, S.J. Hollingsworth, The impact of germline mutations on targeted therapy, The Journal of Pathology 232(2) (2014) 230-243.

Q3: Reviewer #1: Please write full name of "PARP" in line 279 and then write "PARP" inside parentheses alongside it.
A3: Many thanks to the reviewer's comments. **We have revised the manuscript according to the reviewer's suggestions by adding the full name of PARP "poly (ADP-ribose) polymerase (PARP)" and added it in the abbreviations. [Page 9, line 283 and page 15, line 464]**

Q4: Reviewer #1: Please write full name of "CCND2" in line 279 and then write "CCND2" inside parentheses alongside it.
A4: Many thanks to the reviewer's suggestions. **We have made the correction according to the comment by adding the full name of CCND2, "Cyclin-D2 (CCND2)" and added it in the abbreviations. [Page 9, line 284; page 15, line 440]**

Major Comments:

Q1~Q2: Reviewer #1: You should allude to the "statistical test(s)" that you performed in the study at "2.8. Statistical analyses" section. You should allude to the statistical "significance level" at "2.8. Statistical analyses" section.
A1 and A2: Thanks for the comments. We have revised the manuscript according to the reviewer's suggestions**. Over-representation analysis (ORA) was used to prioritize the genes in Knockout Mouse Phenotype, PPI network and Molecular Pathway. The *FDR* of <0.05 was determined to indicate the statistical significance. A hypergeometric test was used to perform an enrichment analysis for PID; the criterion for significance was *p*-value < 0.05. [Page 7, lines 213-216]**

Q3: Reviewer #1: "Literature review" regarding the "SSRIs drugs" at "Discussion" section is not sufficient, please correct the issue.

A3: Many thanks for the reviewer's comments. We have added more information in the part of the discussion for the SSRIs drugs. The revised sentences are as below:

Additionally, CMap analysis also showed that the SSRIs drugs, sertraline, fluoxetine and amitriptyline, may be promising candidates for repurposing to CRC. SSRI drugs have been widely used for the treatment of depression and anxiety [53]. Recently, several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [37-40]. Stopper et al., reported that fluoxetine may reduce tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage in tumor metabolism [38]. A following investigation also supported the anti-cancer activity of sertraline in the two human colon cancer cell lines (HT29 and LS1034) [40]. [Page 12-13, lines 367-374]

Q4: Reviewer #1: You should explain the "limitations" of the study at "Discussion" section.
A4: We sincerely thank the reviewer's suggestions. **The limitation of this study has been added in the manuscript. The sentences are as following**: **However, our approaches have some limitations. The target genes that we identified are not all in pharmacological activities, therefore these might potentially miss the target of the drugs (undruggable), our analysis showed, only 166 drugs (14.98%) of the 1108 protein coding genes are druggable (genetic driven druggable). The comprehensive analysis on the druggability of genes performed by Finan et al. also revealed that only 4479 (22%) among 20,300 protein coding genes are druggable [61]. Moreover, functional studies of the biology of these risk genes and the genes**

targeted by these drugs still require further investigation to ascertain the role of drug target genes. [Page 13, lines 394-401]

Q5: Reviewer #1: You should clear that "All authors read and approved the final manuscript." at "Author Contributions" section.
A5: Thank you for your suggestion, **we added the sentence as suggested by the reviewer. All authors have read and approved of the manuscript and have made significant contributions to this study. [Page 14, lines 418-419]**

Q6: Reviewer #1: You should allocate some sections to "Medical ethics" and "data availability" after "Declaration of Competing Interest" section and explain related issues at the sections.
A6: Thank you for your suggestion. **This study used a public available dataset, no individual data/ sample are used in our study. In addition, we added the Data Availability section after the "Declaration of Competing Interest"**. **[Page 15, lines 424-431]**

Q7: Reviewer #1: References; 9, 24, 48 and 56 are not up-to-date, please substitute the references with up-to-date references. (If possible)
A7: We sincerely thank the reviewer's comments. **We changed the references to be up-to-date according to the reviewer's suggestion. Reference no**: **[9] [Page 3, line 83], [26] [Page 6, line 159], [48] [Page 11, line 339], [56] [Page 12, line 353**].

[9]. H.A. Ghofrani, I.H. Osterloh, F. Grimminger, Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond, Nature reviews. Drug discovery 5(8) (2006) 689-702.
[26]. N. Parvaneh, J.L. Casanova, L.D. Notarangelo*, et al.*, Primary immunodeficiencies: a rapidly evolving story, The Journal of allergy and clinical immunology 131(2) (2013) 314-23.
[48]. J.Y. Wang, J. Sun, M.Y. Huang*, et al.*, STIM1 overexpression promotes colorectal cancer progression, cell motility and COX-2 expression, Oncogene 34(33) (2015) 4358-67.
[56]. A.V. Smirnova, L.B. Lazebnik, I.E. Trubitsina*, et al.*, [Antiproliferative activity of diclofenac at tumor cell cultures], Eksperimental'naia i klinicheskaia gastroenterologiia = Experimental & clinical gastroenterology (5) (2012) 66-9.

Q8: Reviewer #1: Please write "all" the references with a constant style at "References" section.
A8: We are very grateful to the reviewer's suggestion. **We did it as suggested by the reviewer.**

Reviewer 2: In this paper by Irham et al, the authors provide a detailed workflow that harnesses existing genomic data and genetic mapping as a potential method for drug discovery, specifically in the context of colorectal cancer. The authors have done a good job of describing their rationale and results in a detailed manner. Some minor comments are below:

Q1: Reviewer #2: Provide the full form of terms used in the table at the bottom of the respective table itself and in the legend for the figures. For instance, for Table 1, add a row at the end and define the abbreviations used such as GWAS, WES etc.
A1: We thank the reviewer for the suggestions. **We have provided the full name and corrected the tables and legends according to the reviewer's comments.**

Q2: Reviewer #2. Legends for some figures are very brief, a few words in some cases. Need to be much more detailed. Eg: Figs. 3 and 4.
A2: We thank the reviewer's comments. **In the revised manuscript, we have added more detailed explanations in the figure legends of Figure 2 and Figure 3 as suggested. [In the current manuscript, Figure 3 was changed to be Figure 2 and Figure 4 to be Figure 3 due to the removal of Figure 2]. The revised sentences are as below:**

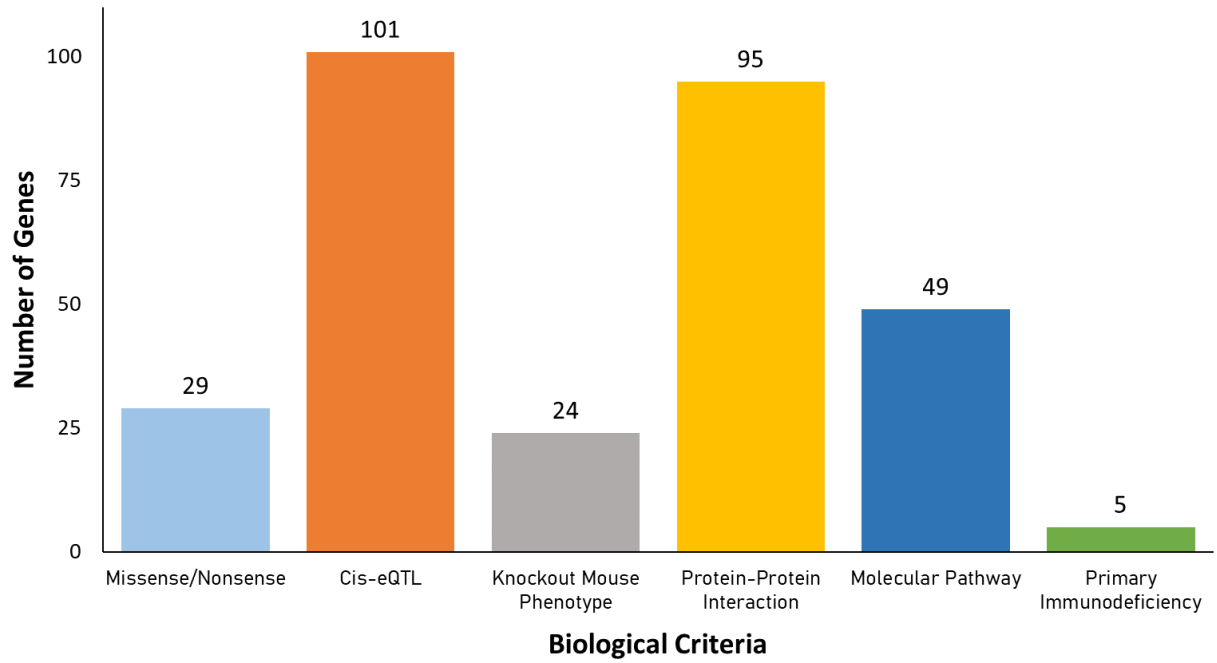


Figure 2. The number of genes satisfying each criterion. **The figure, from left- to the right-side, showed 29 genes which were missense or nonsense variants (blue), 101 genes have cis-expression quantitative trait loci (cis-eQTL) effects on whole blood or colon (orange), 24 genes with knockout mouse phenotype (grey), 95 genes with protein-protein interaction (yellow), 49 CRC-associated genes in molecular pathways (light blue) and five genes overlap in primary immunodeficiency (green).**

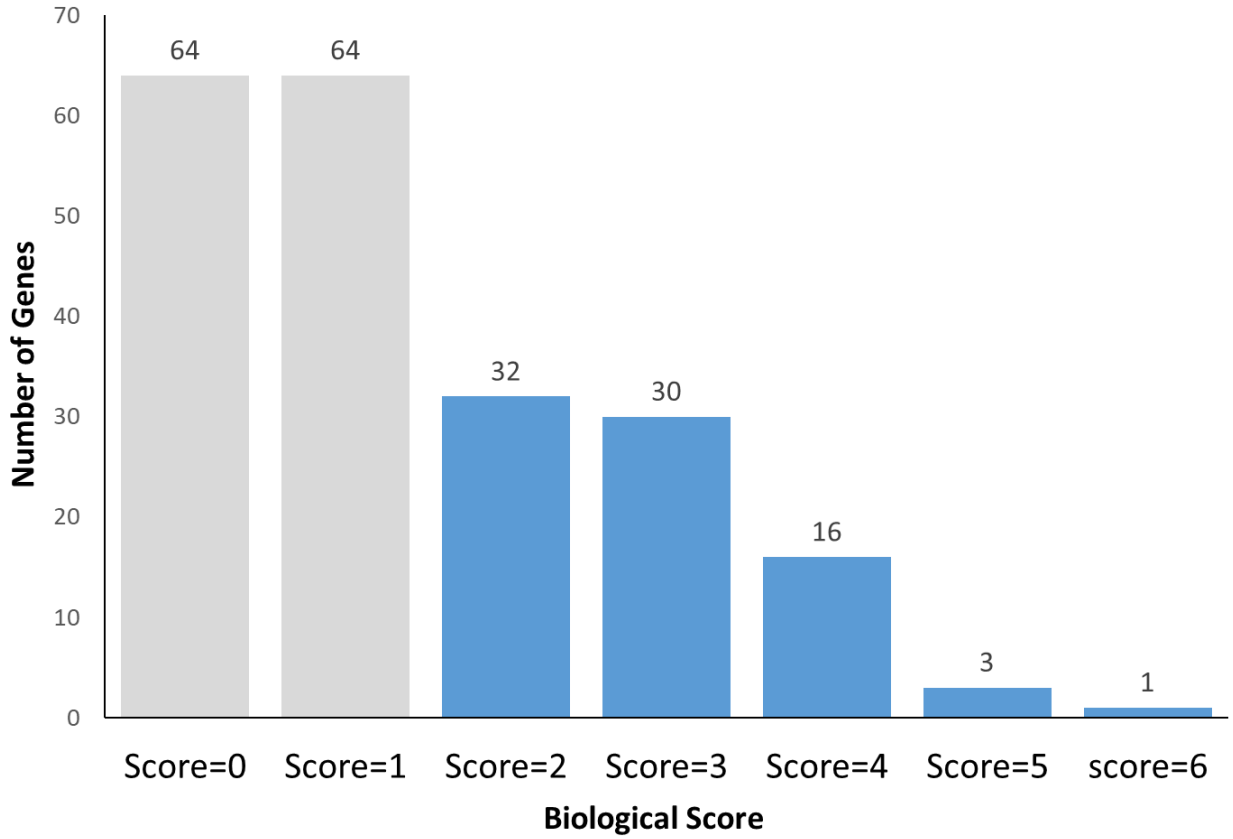


Figure 3. Distribution of colorectal cancer (CRC)-related gene scores. **The figure, from left- to the right-side, showed number of genes with score 0~6. Each gene was assigned one point for each functional annotation. Genes with a score of 0 and 1 have the same numbers, namely 64 genes. Genes with a score of 2 were 32 while those with a score of 3 were 30. There were 16 genes with a score of 4, 3 genes with a score of 5, and 1 gene with a score of 6. After compiling the scores, eighty-two genes with a score of ≥2, which were categorized as 'biological CRC genes'.**

Q3: Reviewer #2. Abbreviations list is missing key definitions for eg: WES
A3: Many thanks to the reviewer's comments. **As suggested, the abbreviations list has been added in the manuscript.**

Reviewer 3: In the manuscript, the authors obtained potential colorectal cancer targets by text mining, HaploReg v4.1 and STRING database, then analyzed drug repurposing for colorectal cancer by Therapeutic Target Database, ClinicalTrials.gov and CMap database. After reviewing this manuscript, I do have some concerns about the research.

Q1: Reviewer #3. In the work of prioritizing genes based on six functional annotation criteria, why did the authors choose these six biological functions as filter criteria? And, why did the authors set the genes with a score of >= 2 as "biological CRC risk genes"? How did the authors get the threshold of the score >= 2?

A1: We appreciate this comment. It is a very important question. In the present study, we prioritized the genes disease and colorectal cancer (CRC) genetics driven genomic drug repurposing for CRC. We hypothesized that CRC genetic variants prioritization using six functional annotations will enable us to translate the risk genes to meaningful insights on CRC pathogenesis. We first mapped the variants onto the corresponding genes with missense/nonsense mutations as one of the non-synonymous changes in the single base substitution of a different amino acid in the resulting protein. We utilized this annotation with the knowledge that functional rules of variants affect protein expression. Furthermore, we leveraged the fact that the expression quantitative trait loci (eQTL) are regions harboring nucleotides correlated with alterations in gene expression. Therefore, the variants may cause changes in gene expression in the direction of the tissues involved (i.e., our analyses focused on the whole blood and colon). If the identified variants cause an upregulation of gene *X,* leading to an increased risk of a disease, then an inhibitor of its protein product may be considered a repositioning candidate. In addition, we applied protein-protein interactions (PPIs) to understand relationships between diseases and biological protein networks. If the genes involved in the biological protein networks are related in CRC pathogenesis, then it is important to inhibit the protein. The genes implicated in knockout mouse phenotype and Kyoto Encyclopedia of Genes and Genomes (KEGG) was also applied to determine the type of molecular pathways enriched on the CRC-associated gene list and the genes involved. The last annotation is the Primary immuno-deficiency (PID) diseases which are innate immune diseases reported to be associated with cancer. Genes overlapping with the PID play a causal role in CRC pathogenesis. It is important to consider the CRC causal relationship and the drug target genes for CRC disease. In addition, these functional annotations have been validated by Yukinori Okada *et al* to prioritize the most likely causal gene relationships with Rheumatoid Arthritis and to find its candidate drugs. According to our analyses, we set the threshold of a biological score >= 2 to find a much higher number of genes as biological CRC genes and candidates of CRC drug targets. Our study showed that the higher the threshold of biological score applied, the smaller the number of biological genes identified, limiting the number of drug targets we could observe. (i.e., we found 4 biological CRC genes for threshold score >=*5,* 20 biological CRC genes for threshold score >=*4,* 50 biological CRC genes for threshold score >=*3* and 82 biological CRC genes for threshold score >= *2*) [The number of genes with all biological scores can be found in **Table S2**]. The more biological CRC genes we find, the more candidate drug targets for CRC drug repurposing can be identified. Unfortunately, the drug target genes that we identified are not all in pharmacological activities, therefore these might potentially miss the target of the drugs (undruggable). Our analysis showed only 166 drugs (14.98%) of the 1108 protein coding genes are druggable (genetic driven druggable) (**Table S4 and Table S6**). Furthermore, we found the threshold score >= 2 from the number of functional

annotations ranged from 0-6, in which each gene was assigned one point for each annotation. Those genes with one functional annotation were awarded one point (score) and those genes with a score >= *2* were classified as "biological CRC genes".

Reference:

Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506(7488):376-381.

Q2: Reviewer #3: In the work of CMap analyses, why did the authors choose capecitabine as the comparison counterpart? Why not other drugs?

A2: Thank you for your comments. **Capecitabine is the standard treatment for CRC. Capecitabine currently is an effective drug for CRC according to the National Comprehensive Cancer Network (NCCN) guideline** [**Stated in page 7, lines 201-202**].

Q3: Reviewer #3. The authors only used existing researches to validate the results of the study, which is not very persuasive. If the experimental verification of the selected genes that are not in existing researches is done, this research will be more convincing and valuable.

A3: Thanks for your comment. It is a very important point. The validation from existing results was necessary and important to ensure whether our drug candidates produce the desired interaction intended in the study, any undesired side effects, or ineffective effects. Along with that, we strongly agree that if we could verify genes which are not in existing research, the result would be more valuable. However, our present work focuses on narrowing down the candidate drugs through leveraging existing genomic data and genetic mapping as a potential method to guide drug repurposing for CRC. We admit this methodology has some limitations, as we added in the last paragraph of the discussion as part of the limitations of this study. **"Moreover, functional study of the biology of these risk genes and the genes targeted by these drugs still require further investigation to ascertain the role of drug target genes".** [**Page 13, lines 394-396]**

Q4: Reviewer #3. What are the innovations of this study?

A4: We sincerely thank the reviewer for taking the time to review our work. Our study attempted to integrate the genetics of colorectal cancer (CRC) with diverse biological resources to aid CRC drug repurposing. We utilized functional and bioinformatics annotations of CRC risk variants and integrated current CRC genetic findings with the complete catalog of approved drugs for CRC and other diseases. Our findings noted that some drugs may be promising candidates for repurposing to CRC and have not been previously reported using the bioinformatics and functional annotation approach (i.e. Selective Serotonin Reuptake Inhibitors (SSRI) (sertraline, fluoxetine and amitriptyline) target SLC6A4). For example, the original indication of SSRI drugs are for the treatment of depression and anxiety [1]. To validate these findings, we proceeded to search previous existing literature for information on these candidates. Several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [2-5]. Stopper et al., reported that fluoxetine may reduce tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage in tumor metabolism [3]. A following investigation also supported the anti-cancer activity of sertraline in two human colon cancer cell lines (HT29 and LS1034) [5]. We also identified another drug potentially useful for treating CRC is vitamin D. Vitamin D deficiency may affect various

human cancers [6]. Vitamin D has been shown to protect from CRC tumorigenesis by binding long-chain fatty acids and bile acids in the small intestine and protecting colonic epithelial cells from mutagens [7]. Notably, Vitamin D-related drugs are currently being evaluated in CRC: cholecalciferol is in a phase 2 trial [NCT01074216]; and paricalcitol [NCT01197664] is in phase 1 trials. In addition, several non-steroidal anti-inflammatory drugs (NSAID) drugs (i.e. Ibuprofen, naproxen and celecoxib) also might be promising for CRC drug repurposing. Validation an in vitro study on adenocarcinoma gastric cells, the antitumor activity of ibuprofen was attributed to the induction of apoptosis and the reduction of cell proliferation, which attenuated angiogenesis [8]. According to an in vivo study showed that the combination of naproxen and atorvastatin significantly inhibited growth of colonic adenocarcinomas [9]. Furthermore, a study of celecoxib as an adjuvant chemotherapy is ongoing in patients with metastatic CRC (phase 4) [NCT03645187]. Our finding suggests several drug target genes that we identified in this study have not been reported yet in previous study based on the bioinformatics and functional annotations approach, furthermore, the biological plausibility of repurposing cancer drugs from non-cancer drugs or other drugs to treatment of CRC.

References:

[1] J. Ramsberg, C. Asseburg, M. Henriksson, Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model, PloS one 7(8) (2012) e42003-e42003.
[2] H. Arimochi, K. Morita, Characterization of cytotoxic actions of tricyclic antidepressants on human HT29 colon carcinoma cells, European journal of pharmacology 541(1-2) (2006) 17-23.
[3] H. Stopper, S.B. Garcia, A.M. Waaga-Gasser*, et al.*, Antidepressant fluoxetine and its potential against colon tumors, World J Gastrointest Oncol 6(1) (2014) 11-21.
[4] W.J. Jang, S.K. Jung, T.T.L. Vo*, et al.*, Anticancer activity of paroxetine in human colon cancer cells: Involvement of MET and ERBB3, Journal of cellular and molecular medicine 23(2) (2019) 1106-1115.
[5] I. Gil-Ad, A. Zolokov, L. Lomnitski*, et al.*, Evaluation of the potential anti-cancer activity of the antidepressant sertraline in human colon cancer cell lines and in colorectal cancer-xenografted mice, International journal of oncology 33(2) (2008) 277-86.
[6] J. Sun, The Role of Vitamin D and Vitamin D Receptors in Colon Cancer, Clin Transl Gastroenterol 8(6) (2017) e103-e103.
[7] S.A. Lamprecht, M. Lipkin, Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis, Annals of the New York Academy of Sciences 952 (2001) 73-87.
[8] H. Akrami, S. Aminzadeh, H. Fallahi, Inhibitory effect of ibuprofen on tumor survival and angiogenesis in gastric cancer cell, Tumour biology 36(5) (2015) 3237-43.
[9] N. Suh, B.S. Reddy, A. DeCastro*, et al.*, Combination of atorvastatin with sulindac or naproxen profoundly inhibits colonic adenocarcinomas by suppressing the p65/beta-catenin/cyclin D1 signaling pathway in rats, Cancer prevention research (Philadelphia, Pa.) 4(11) (2011) 1895-902.

Q5: Reviewer #3. Page 10, lines 298, 299 and 310: The full name of the abbreviation "mCRC", not in abbreviation list, should be given, or it is difficult to understand the meaning.
A5: We sincerely thank the reviewer for their suggestion. **We already revised this in the manuscript according to the reviewer's suggestions. [Page 10, lines 301, 302 and page 11, line 314]**

Bukti Accepted Artikel yang dikirim ke Corresponding Author:

Decision on submission to Pharmacological Research - [EMID:18edffdebc565bb3]

收件匣

Pharmacological Re... 上午 5:40

寄給我

CC: deadoc80@gmail.com

Manuscript Number: YPHRS-D-20-01156R1
Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer

Dear Prof. Chang,

I am pleased to inform you that your paper has been accepted for publication. Now that your manuscript has been accepted for publication it will proceed to copy-editing and production.

Thank you for submitting your work to Pharmacological Research. We hope you consider us again for future submissions.

Kind regards,

Pharmacological Research

Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer

--Manuscript Draft--

Manuscript Number:	YPHRS-D-20-01156R1
Article Type:	Research Paper
Keywords:	Drug discovery; drug repurposed; data mining; colorectal cancer; bioinformatics
Corresponding Author:	Wei-Chiao Chang, PhD TAIWAN
First Author:	Ialu Muhammad Irham, M.Sc
Order of Authors:	Ialu Muhammad Irham, M.Sc
	Henry Sung-Ching Wong
	Wirawan Adikusuma, M.Sc
	Wan-Hsuan Chou, M.Sc
	Wan-Chen Huang, PhD
	Wei-Chiao Chang, PhD
	Eko Mugiyanto, M.Sc
Abstract:	Even though many genetic risk loci for human diseases have been identified and comprehensively cataloged, strategies to guide clinical research by integrating the extensive results of genetic studies and biological resources are still limited. Moreover, integrative analyses that provide novel insights into disease biology are expected to be especially useful for drug discovery. Herein, we use text mining of genetic studies on colorectal cancer (CRC) and assign biological annotations to identified risk genes in order to discover novel drug targets and potential drugs for repurposing. Risk genes for CRC were obtained from PubMed text mining, and for each gene, six functional and bioinformatic annotations were analyzed. The annotations include missense mutations, expression quantitative trait loci (eQTL), molecular pathway analyses, protein-protein interactions (PPIs), genetic overlap with knockout mouse phenotypes, and primary immunodeficiency. We then prioritized biological risk candidate genes according to a scoring system for the six functional annotations. Each functional annotation was assigned one point, and those genes with a score ≥2 were designated “biological CRC risk genes”. Using this method, we revealed 82 biological CRC risk genes, which mapped to 128 genes in an expanded PPI network. Further utilizing DrugBank and the Therapeutic Target Database (TTD), we found 21 genes in our list that are targeted by 166 candidate drugs. Based on data from ClinicalTrials. gov , we found our list contains four known target genes with six drugs approved for CRC treatment, as well as three known target genes with nine drugs under preclinical investigation for CRC. Additionally, 12 genes are targeted by 32 drugs approved for other indications, which can possibly be repurposed for CRC treatment. Finally, our analysis from Connectivity Map (CMap) showed that 18 of the 41 drugs under clinical and preclinical investigation have high potential to be useful for CRC.
Suggested Reviewers:	

August 26, 2020

Dear Editors,

Please find our attached manuscript entitled "**Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer**," which we are submitting for consideration for publication as an Original Research article in ***Pharmacological Research*** (**YPHRS-D-20-01156**). We are thankful for your kind suggestions regarding our manuscript. Here, we are sending our revised manuscript in accordance with the comments given by the three reviewers. We have read through all the reviewers' suggestions very carefully, and made the necessary revisions based on these comments, as detailed below in a point-by-point format. The revised sections are highlighted in red. Finally, we would like to thank you once again for giving us the opportunity to improve our manuscript. We very much hope that these revisions are adequate. We appreciate your assistance and are looking forward to hearing from you.

Sincerely yours,

Wei-Chiao Chang (D.Phil.; Oxon)
Professor, Department of Clinical Pharmacy,
Taipei Medical University, Taiwan
250 Wu-Hsing Street, Taipei 110, Taiwan

Table 1: ***Author Checklist for Original Articles to be submitted to Pharmacological Research***

Questions	***reply***		
	Yes	***No***	***Not applicable***
Formatting - The submission will automatically be rejected if the first six questions are not marked "yes" (questions 1-6) or "not applicable" (limited to questions 3-6**)**			
1. Are all tables and figures numbered and appropriately titled with descriptive legends that permit stand-alone interpretation?	√		
2. Are all data shown on figures and tables mentioned in the text of the Results section?	√		
3. Are the whole un-cropped images of the original western blots from which figures have been derived shown as supplemental figures?	√		
4. In case of human studies, has specific mention been made of the study compliance with the regulations of the country(ies) in which the study was carried out ?	√		
5. In case of human studies, has the study been registered on an accessible international registry/database of clinical trials (e.g. EudraCT, ClinicalTrials.gov, ChiCTR, ANZCTR, JPRN and the like)	√		
6. In case of studies on animals, is there a statement indicating compliance with regulations on the ethical treatment of animals including identification of the institutional committee that approved the experiments?	√		
Introduction			
7. Is there a clear statement with background describing the hypothesis being tested by this study?	√		
8. Are the primary endpoints clearly described?	√		
Materials and Methods			
9. Are the sources of all materials clearly indicated?	√		
10. Is(are) the chemical structure(s) of any new compound(s) presented as a figure in the manuscript or referenced in the literature?	√		
11. Are the source(s) and passage number of cell lines indicated?			√
12. Are the source, catalogue number and lot for commercial antibodies indicated?			√
13. Are the species, strain, sex, weight and source of the animal subjects provided?			√
14. Is the rationale provided for the selection of concentrations, doses, route and frequency of compound administration?			√
15. Are quantified results (*e.g.* IC_{50} and/or EC_{50} values) of concentration- and dose-response experiments included in the report?			√
16. Is the method of anaesthesia described?			√
17. Are all group sizes approximately the same?			√
18. Were the criteria used for excluding any data from analysis determined prospectively and clearly stated?			√
19. Was the investigator responsible for data analysis blinded to which samples/animals represent treatments and controls?			√
20. Is the exact sample size (n) for each experimental group/condition clearly indicated in the text and/or on the tables and figures?			√

21. Are the reported data displayed as the mean +/- an estimate of variability (SD, SEM) of three or more independent experimental replications?			√
22. Is the number of replications used to generate an individual data point in each of the independent experiments clearly indicated?			√
23. Were the statistical tests employed to analyse the primary endpoints predetermined as part of the experimental design?			√
24. Is the threshold for statistical significance (P value) clearly indicated?	√		
25. Were the data normalised?			√
26. Were post-hoc tests used to assess the statistical significance among means?			√
27. Was the study exploratory rather than hypothesis-driven?			√
28. Were human tissues or fluids used in this study?			√
Results			
29. If western blots are shown, are the following included: i) appropriate loading controls for each western blot, ii) replication data, iii) quantification, and iv) the results of a statistical analysis?			√
30. Were MIQE guidelines followed in the quantitative analysis and presentation of PCR and RT-PCR findings?			√
31. Was a reference standard (positive or negative controls) included in the study to validate the experiment?			√
Discussion			
32. Are all the findings considered within the context of the hypothesis presented in the Introduction?	√		
33. Are the primary conclusions and their implications clearly stated?	√		
34. Are any secondary endpoints reported and are these sufficiently powered for appropriate statistical analysis?	√		
35. Are the limitations of the current study or alternative interpretations of the findings clearly stated?	√		
For Meta-analyses only			
36. Were the PRISMA reporting guidelines and checklist followed in the case of meta-analyses on randomised controlled trials?			
37. Were the MOOSE reporting guidelines followed in the case of meta-analyses on observational studies in epidemiology?			
38. Was the protocol submitted into the PROSPERO International prospective register of systematic reviews and a registration number obtained?			
Conflict of Interest/Financial Support			
39. Is the conflict of interest statement is included in the manuscript?	√		
40. Are all organisations providing funding for this work listed in the Acknowledgments?	√		

Feedback/suggestions on the checklist by the author

August 26, 2020

Dear Editors,

Please find our attached manuscript entitled "**Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer**," which we are submitting for consideration for publication as an Original Research article in ***Pharmacological Research*** (**YPHRS-D-20-01156**). We are thankful for your kind suggestions regarding our manuscript. Here, we are sending our revised manuscript in accordance with the comments given by the three reviewers. We have read through all the reviewers' suggestions very carefully, and made the necessary revisions based on these comments, as detailed below in a point-by-point format. The revised sections are highlighted in red. Finally, we would like to thank you once again for giving us the opportunity to improve our manuscript. We very much hope that these revisions are adequate. We appreciate your assistance and are looking forward to hearing from you.

Sincerely yours,

Wei-Chiao Chang (D.Phil.; Oxon)
Professor, Department of Clinical Pharmacy,
Taipei Medical University, Taiwan
250 Wu-Hsing Street, Taipei 110, Taiwan

Reviewer Comment:

Reviewer 1: Manuscript Title: Integration of Genetic Variants and Gene Network for Drug Repositioning in Colorectal Cancer (Lalu Muhammad Irham1,2,3, Henry Sung-Ching Wong1,2, Wan-Hsuan Chou1,2, Wirawan Adikusuma1,2,4, Eko Mugiyanto1,2,5, Wan-Chen Huang6*, Wei-Chiao Chang1,2,7,8,9*)

General Comments:

The manuscript entitled "Integration of Genetic Variants and Gene Network for Drug Repositioning in Colorectal Cancer" is comprehensive study of the drug repurposing in colorectal cancer generally. Title of article is appropriate. Abstract is good arranged. Introduction is good organized with sufficient literature review. Methods are evidence based. Results show findings clearly. Conclusion is logical and appropriate. References are related to the issue. I think the figures, tables and highlights are important and help to better understanding of the subject although I have some corrections to improve the article value that I describe in "Minor & Major comments".

Answer: We thank the reviewer's comments

Minor Comments:

Q1: Reviewer #1: Please change "Repositioning" to "Repurposing" in "title" of article.
A1: We thank the reviewer's comments. **Yes, we have revised the word: "Repositioning" to "Repurposing" in the title of the manuscript according to the reviewer's suggestions. As shown in the title "Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer". [Page 1, lines 3-4]**

Q2: Reviewer #1: Please add related reference(s) after "drugs." in line 103.
A2: Thanks for the reviewer's suggestions. **We have revised the manuscript by adding two references [16 and 17]. These two references are shown as below, and we added them accordingly in the manuscript. [Page 4, line 103]**

[16] E.M. Stoffel, E. Koeppe, J. Everett, *et al.*, Germline Genetic Features of Young Individuals With Colorectal Cancer, Gastroenterology 154(4) (2018) 897-905.e1.
[17] S.A. Smith, T. French, S.J. Hollingsworth, The impact of germline mutations on targeted therapy, The Journal of Pathology 232(2) (2014) 230-243.

Q3: Reviewer #1: Please write full name of "PARP" in line 279 and then write "PARP" inside parentheses alongside it.
A3: Many thanks to the reviewer's comments. **We have revised the manuscript according to the reviewer's suggestions by adding the full name of PARP "poly (ADP-ribose) polymerase (PARP)" and added it in the abbreviations. [Page 9, line 283 and page 15, line 464]**

Q4: Reviewer #1: Please write full name of "CCND2" in line 279 and then write "CCND2" inside parentheses alongside it.
A4: Many thanks to the reviewer's suggestions. **We have made the correction according to the comment by adding the full name of CCND2, "Cyclin-D2 (CCND2)" and added it in the abbreviations. [Page 9, line 284; page 15, line 440]**

Major Comments:

Q1~Q2: Reviewer #1: You should allude to the "statistical test(s)" that you performed in the study at "2.8. Statistical analyses" section. You should allude to the statistical "significance level" at "2.8. Statistical analyses" section.
A1 and A2: Thanks for the comments. We have revised the manuscript according to the reviewer's suggestions**. Over-representation analysis (ORA) was used to prioritize the genes in Knockout Mouse Phenotype, PPI network and Molecular Pathway. The *FDR* of <0.05 was determined to indicate the statistical significance. A hypergeometric test was used to perform an enrichment analysis for PID; the criterion for significance was *p*-value < 0.05. [Page 7, lines 213-216]**

Q3: Reviewer #1: "Literature review" regarding the "SSRIs drugs" at "Discussion" section is not sufficient, please correct the issue.

A3: Many thanks for the reviewer's comments. We have added more information in the part of the discussion for the SSRIs drugs. The revised sentences are as below:

Additionally, CMap analysis also showed that the SSRIs drugs, sertraline, fluoxetine and amitriptyline, may be promising candidates for repurposing to CRC. SSRI drugs have been widely used for the treatment of depression and anxiety [53]. Recently, several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [37-40]. Stopper et al., reported that fluoxetine may reduce tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage in tumor metabolism [38]. A following investigation also supported the anti-cancer activity of sertraline in the two human colon cancer cell lines (HT29 and LS1034) [40]. [Page 12-13, lines 367-374]

Q4: Reviewer #1: You should explain the "limitations" of the study at "Discussion" section.
A4: We sincerely thank the reviewer's suggestions. **The limitation of this study has been added in the manuscript. The sentences are as following**: **However, our approaches have some limitations. The target genes that we identified are not all in pharmacological activities, therefore these might potentially miss the target of the drugs (undruggable), our analysis showed, only 166 drugs (14.98%) of the 1108 protein coding genes are druggable (genetic driven druggable). The comprehensive analysis on the druggability of genes performed by Finan et al. also revealed that only 4479 (22%) among 20,300 protein coding genes are**

druggable [61]. Moreover, functional studies of the biology of these risk genes and the genes targeted by these drugs still require further investigation to ascertain the role of drug target genes. [Page 13, lines 394-401]

Q5: Reviewer #1: You should clear that "All authors read and approved the final manuscript." at "Author Contributions" section.
A5: Thank you for your suggestion, **we added the sentence as suggested by the reviewer. All authors have read and approved of the manuscript and have made significant contributions to this study. [Page 14, lines 418-419]**

Q6: Reviewer #1: You should allocate some sections to "Medical ethics" and "data availability" after "Declaration of Competing Interest" section and explain related issues at the sections.
A6: Thank you for your suggestion. **This study used a public available dataset, no individual data/ sample are used in our study. In addition, we added the Data Availability section after the "Declaration of Competing Interest"**. **[Page 15, lines 424-431]**

Q7: Reviewer #1: References; 9, 24, 48 and 56 are not up-to-date, please substitute the references with up-to-date references. (If possible)
A7: We sincerely thank the reviewer's comments. **We changed the references to be up-to-date according to the reviewer's suggestion. Reference no**: **[9] [Page 3, line 83], [26] [Page 6, line 159], [48] [Page 11, line 339], [56] [Page 12, line 353**].

[9]. H.A. Ghofrani, I.H. Osterloh, F. Grimminger, Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond, Nature reviews. Drug discovery 5(8) (2006) 689-702.
[26]. N. Parvaneh, J.L. Casanova, L.D. Notarangelo*, et al.*, Primary immunodeficiencies: a rapidly evolving story, The Journal of allergy and clinical immunology 131(2) (2013) 314-23.
[48]. J.Y. Wang, J. Sun, M.Y. Huang*, et al.*, STIM1 overexpression promotes colorectal cancer progression, cell motility and COX-2 expression, Oncogene 34(33) (2015) 4358-67.
[56]. A.V. Smirnova, L.B. Lazebnik, I.E. Trubitsina*, et al.*, [Antiproliferative activity of diclofenac at tumor cell cultures], Eksperimental'naia i klinicheskaia gastroenterologiia = Experimental & clinical gastroenterology (5) (2012) 66-9.

Q8: Reviewer #1: Please write "all" the references with a constant style at "References" section.
A8: We are very grateful to the reviewer's suggestion. **We did it as suggested by the reviewer.**

Reviewer 2: In this paper by Irham et al, the authors provide a detailed workflow that harnesses existing genomic data and genetic mapping as a potential method for drug discovery, specifically in the context of colorectal cancer. The authors have done a good job of describing their rationale and results in a detailed manner. Some minor comments are below:

Q1: Reviewer #2: Provide the full form of terms used in the table at the bottom of the respective table itself and in the legend for the figures. For instance, for Table 1, add a row at the end and define the abbreviations used such as GWAS, WES etc.
A1: We thank the reviewer for the suggestions. **We have provided the full name and corrected the tables and legends according to the reviewer's comments.**

Q2: Reviewer #2. Legends for some figures are very brief, a few words in some cases. Need to be much more detailed. Eg: Figs. 3 and 4.
A2: We thank the reviewer's comments. **In the revised manuscript, we have added more detailed explanations in the figure legends of Figure 2 and Figure 3 as suggested**. **[In the current manuscript, Figure 3 was changed to be Figure 2 and Figure 4 to be Figure 3 due to the removal of Figure 2]. The revised sentences are as below:**

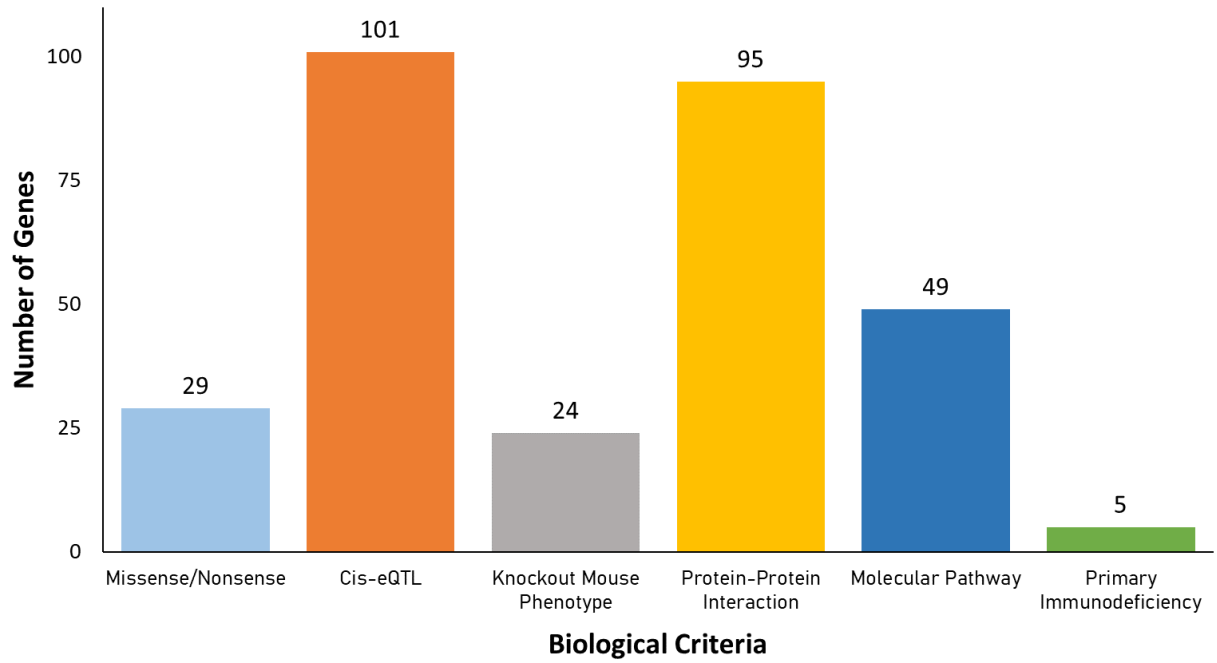


Figure 2. The number of genes satisfying each criterion. **The figure, from left- to the right-side, showed 29 genes which were missense or nonsense variants (blue), 101 genes have cis-expression quantitative trait loci (cis-eQTL) effects on whole blood or colon (orange), 24 genes with knockout mouse phenotype (grey), 95 genes with protein-protein interaction (yellow), 49 CRC-associated genes in molecular pathways (light blue) and five genes overlap in primary immunodeficiency (green).**

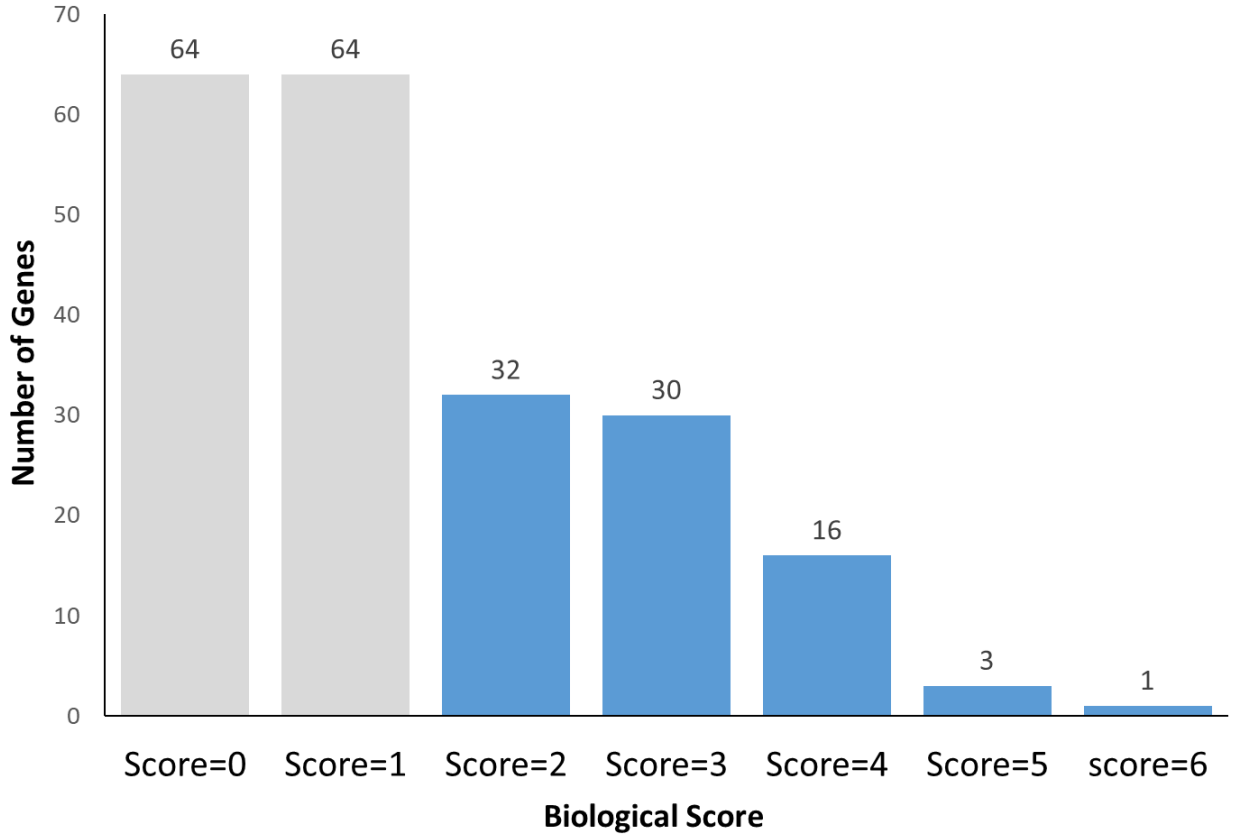


Figure 3. Distribution of colorectal cancer (CRC)-related gene scores. **The figure, from left- to the right-side, showed number of genes with score 0~6. Each gene was assigned one point for each functional annotation. Genes with a score of 0 and 1 have the same numbers, namely 64 genes. Genes with a score of 2 were 32 while those with a score of 3 were 30. There were 16 genes with a score of 4, 3 genes with a score of 5, and 1 gene with a score of 6. After compiling the scores, eighty-two genes with a score of ≥2, which were categorized as 'biological CRC genes'.**

Q3: Reviewer #2. Abbreviations list is missing key definitions for eg: WES
A3: Many thanks to the reviewer's comments. **As suggested**, **the abbreviations list has been added in the manuscript.**

Reviewer 3: In the manuscript, the authors obtained potential colorectal cancer targets by text mining, HaploReg v4.1 and STRING database, then analyzed drug repurposing for colorectal cancer by Therapeutic Target Database, ClinicalTrials.gov and CMap database. After reviewing this manuscript, I do have some concerns about the research.

Q1: Reviewer #3. In the work of prioritizing genes based on six functional annotation criteria, why did the authors choose these six biological functions as filter criteria? And, why did the authors set the genes with a score of >= 2 as "biological CRC risk genes"? How did the authors get the threshold of the score >= 2?

A1: We appreciate this comment. It is a very important question. In the present study, we prioritized the genes disease and colorectal cancer (CRC) genetics driven genomic drug repurposing for CRC. We hypothesized that CRC genetic variants prioritization using six functional annotations will enable us to translate the risk genes to meaningful insights on CRC pathogenesis. We first mapped the variants onto the corresponding genes with missense/nonsense mutations as one of the non-synonymous changes in the single base substitution of a different amino acid in the resulting protein. We utilized this annotation with the knowledge that functional rules of variants affect protein expression. Furthermore, we leveraged the fact that the expression quantitative trait loci (eQTL) are regions harboring nucleotides correlated with alterations in gene expression. Therefore, the variants may cause changes in gene expression in the direction of the tissues involved (i.e., our analyses focused on the whole blood and colon). If the identified variants cause an upregulation of gene *X,* leading to an increased risk of a disease, then an inhibitor of its protein product may be considered a repositioning candidate. In addition, we applied protein-protein interactions (PPIs) to understand relationships between diseases and biological protein networks. If the genes involved in the biological protein networks are related in CRC pathogenesis, then it is important to inhibit the protein. The genes implicated in knockout mouse phenotype and Kyoto Encyclopedia of Genes and Genomes (KEGG) was also applied to determine the type of molecular pathways enriched on the CRC-associated gene list and the genes involved. The last annotation is the Primary immuno-deficiency (PID) diseases which are innate immune diseases reported to be associated with cancer. Genes overlapping with the PID play a causal role in CRC pathogenesis. It is important to consider the CRC causal relationship and the drug target genes for CRC disease. In addition, these functional annotations have been validated by Yukinori Okada *et al* to prioritize the most likely causal gene relationships with Rheumatoid Arthritis and to find its candidate drugs. According to our analyses, we set the threshold of a biological score >= 2 to find a much higher number of genes as biological CRC genes and candidates of CRC drug targets. Our study showed that the higher the threshold of biological score applied, the smaller the number of biological genes identified, limiting the number of drug targets we could observe. (i.e., we found 4 biological CRC genes for threshold score >=*5,* 20 biological CRC genes for threshold score >=*4,* 50 biological CRC genes for threshold score >=*3* and 82 biological CRC genes for threshold score >= *2*) [The number of genes with all biological scores can be found in **Table S2**]. The more biological CRC genes we find, the more candidate drug targets for CRC drug repurposing can be identified. Unfortunately, the drug target genes that we identified are not all in pharmacological activities, therefore these might potentially miss the target of the drugs (undruggable). Our analysis showed only 166 drugs (14.98%) of the 1108 protein coding genes are druggable (genetic driven druggable) (**Table S4 and Table S6**). Furthermore, we found the threshold score >= 2 from the number of functional

annotations ranged from 0-6, in which each gene was assigned one point for each annotation. Those genes with one functional annotation were awarded one point (score) and those genes with a score >= *2* were classified as "biological CRC genes".

Reference:

Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506(7488):376-381.

Q2: Reviewer #3: In the work of CMap analyses, why did the authors choose capecitabine as the comparison counterpart? Why not other drugs?

A2: Thank you for your comments. **Capecitabine is the standard treatment for CRC. Capecitabine currently is an effective drug for CRC according to the National Comprehensive Cancer Network (NCCN) guideline** [**Stated in page 7, lines 201-202**].

Q3: Reviewer #3. The authors only used existing researches to validate the results of the study, which is not very persuasive. If the experimental verification of the selected genes that are not in existing researches is done, this research will be more convincing and valuable.

A3: Thanks for your comment. It is a very important point. The validation from existing results was necessary and important to ensure whether our drug candidates produce the desired interaction intended in the study, any undesired side effects, or ineffective effects. Along with that, we strongly agree that if we could verify genes which are not in existing research, the result would be more valuable. However, our present work focuses on narrowing down the candidate drugs through leveraging existing genomic data and genetic mapping as a potential method to guide drug repurposing for CRC. We admit this methodology has some limitations, as we added in the last paragraph of the discussion as part of the limitations of this study. **"Moreover, functional study of the biology of these risk genes and the genes targeted by these drugs still require further investigation to ascertain the role of drug target genes".** [**Page 13, lines 394-396]**

Q4: Reviewer #3. What are the innovations of this study?

A4: We sincerely thank the reviewer for taking the time to review our work. Our study attempted to integrate the genetics of colorectal cancer (CRC) with diverse biological resources to aid CRC drug repurposing. We utilized functional and bioinformatics annotations of CRC risk variants and integrated current CRC genetic findings with the complete catalog of approved drugs for CRC and other diseases. Our findings noted that some drugs may be promising candidates for repurposing to CRC and have not been previously reported using the bioinformatics and functional annotation approach (i.e. Selective Serotonin Reuptake Inhibitors (SSRI) (sertraline, fluoxetine and amitriptyline) target SLC6A4). For example, the original indication of SSRI drugs are for the treatment of depression and anxiety [1]. To validate these findings, we proceeded to search previous existing literature for information on these candidates. Several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [2-5]. Stopper et al., reported that fluoxetine may reduce tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage in tumor metabolism [3]. A following investigation also supported the anti-cancer activity of sertraline in two human colon cancer cell lines (HT29 and LS1034) [5]. We also identified another drug potentially useful for treating CRC is vitamin D. Vitamin D deficiency may affect various

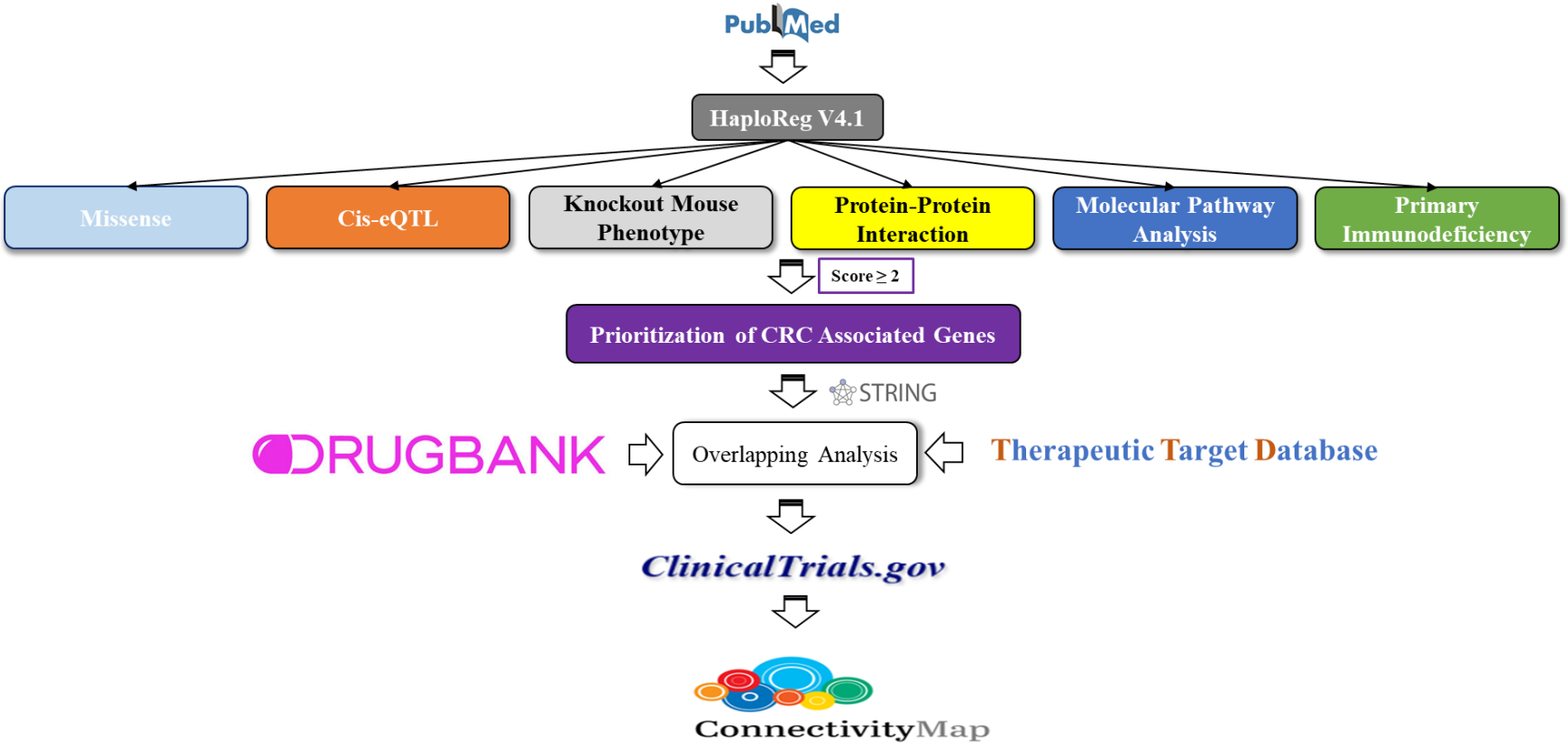
human cancers [6]. Vitamin D has been shown to protect from CRC tumorigenesis by binding long-chain fatty acids and bile acids in the small intestine and protecting colonic epithelial cells from mutagens [7]. Notably, Vitamin D-related drugs are currently being evaluated in CRC: cholecalciferol is in a phase 2 trial [NCT01074216]; and paricalcitol [NCT01197664] is in phase 1 trials. In addition, several non-steroidal anti-inflammatory drugs (NSAID) drugs (i.e. Ibuprofen, naproxen and celecoxib) also might be promising for CRC drug repurposing. Validation an in vitro study on adenocarcinoma gastric cells, the antitumor activity of ibuprofen was attributed to the induction of apoptosis and the reduction of cell proliferation, which attenuated angiogenesis [8]. According to an in vivo study showed that the combination of naproxen and atorvastatin significantly inhibited growth of colonic adenocarcinomas [9]. Furthermore, a study of celecoxib as an adjuvant chemotherapy is ongoing in patients with metastatic CRC (phase 4) [NCT03645187]. Our finding suggests several drug target genes that we identified in this study have not been reported yet in previous study based on the bioinformatics and functional annotations approach, furthermore, the biological plausibility of repurposing cancer drugs from non-cancer drugs or other drugs to treatment of CRC.

References:

[1] J. Ramsberg, C. Asseburg, M. Henriksson, Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model, PloS one 7(8) (2012) e42003-e42003.
[2] H. Arimochi, K. Morita, Characterization of cytotoxic actions of tricyclic antidepressants on human HT29 colon carcinoma cells, European journal of pharmacology 541(1-2) (2006) 17-23.
[3] H. Stopper, S.B. Garcia, A.M. Waaga-Gasser*, et al.*, Antidepressant fluoxetine and its potential against colon tumors, World J Gastrointest Oncol 6(1) (2014) 11-21.
[4] W.J. Jang, S.K. Jung, T.T.L. Vo*, et al.*, Anticancer activity of paroxetine in human colon cancer cells: Involvement of MET and ERBB3, Journal of cellular and molecular medicine 23(2) (2019) 1106-1115.
[5] I. Gil-Ad, A. Zolokov, L. Lomnitski*, et al.*, Evaluation of the potential anti-cancer activity of the antidepressant sertraline in human colon cancer cell lines and in colorectal cancer-xenografted mice, International journal of oncology 33(2) (2008) 277-86.
[6] J. Sun, The Role of Vitamin D and Vitamin D Receptors in Colon Cancer, Clin Transl Gastroenterol 8(6) (2017) e103-e103.
[7] S.A. Lamprecht, M. Lipkin, Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis, Annals of the New York Academy of Sciences 952 (2001) 73-87.
[8] H. Akrami, S. Aminzadeh, H. Fallahi, Inhibitory effect of ibuprofen on tumor survival and angiogenesis in gastric cancer cell, Tumour biology 36(5) (2015) 3237-43.
[9] N. Suh, B.S. Reddy, A. DeCastro*, et al.*, Combination of atorvastatin with sulindac or naproxen profoundly inhibits colonic adenocarcinomas by suppressing the p65/beta-catenin/cyclin D1 signaling pathway in rats, Cancer prevention research (Philadelphia, Pa.) 4(11) (2011) 1895-902.

Q5: Reviewer #3. Page 10, lines 298, 299 and 310: The full name of the abbreviation "mCRC", not in abbreviation list, should be given, or it is difficult to understand the meaning.
A5: We sincerely thank the reviewer for their suggestion. **We already revised this in the manuscript according to the reviewer's suggestions. [Page 10, lines 301, 302 and page 11, line 314]**

Graphical Abstract. Integrative analyses of genetics-driven genomic drug repurposing for colorectal cancer (CRC) by utilizing multiple databases. CRC-associated Single Nucleotide Polymorphisms (SNPs) were identified by searching PubMed literature. The list was expanded with HaploReg v4.1 to find additional CRC-associated genes. Six functional annotations criteria were used to generate a biological prioritization score: (1) CRC risk missense variant; (2) Cis-eQTL; (3) Knockout mouse phenotype; (4) Protein-Protein Interaction; (5) Molecular pathway analysis; and (6) Primary immunodeficiency. Those genes with one functional annotation were awarded one point (score) and those genes with a score ≥2 were classified as "biological CRC genes". The biological CRC genes were utilized in further analyses, including the use of the STRING database to expand the list of candidate genes. The extent of overlap with target genes for approved CRC drugs was also assessed using the DrugBank database and Therapeutic Target Database (TTD). Drugs in clinical trials were identified using ClinicalTrials.gov. Finally, the CMap database was used to prioritize the most promising drugs for CRC treatment.



Original Article

Integration of Genetic Variants and Gene Network for Drug Repurposing in Colorectal Cancer

Lalu Muhammad Irham[1,2,3], Henry Sung-Ching Wong[1,2], Wan-Hsuan Chou[1,2], Wirawan Adikusuma[1,2,4], Eko Mugiyanto[1,2,5], Wan-Chen Huang[6]*, Wei-Chiao Chang[1,2,7,8,9]*

[1]*Department of Clinical Pharmacy, School of Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, 11031, Taiwan*

[2]*Master Program for Clinical Pharmacogenomics and Pharmacoproteomics, School of Pharmacy, Taipei Medical University, Taipei, 11031, Taiwan*

[3]*Faculty of Pharmacy, University of Ahmad Dahlan, Yogyakarta, 55164, Indonesia*

[4]*Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, 83127, Indonesia*

[5]*Department of pharmacy, Faculty of Health Science, University of Muhammadiyah Pekajangan Pekalongan, 51172, Indonesia*

[6]*Institute of Cellular and Organismic Biology, Academia Sinica, Taipei City 115, Taiwan*

[7]*Integrative Research Center for Critical Care, Wan fang Hospital, Taipei Medical University-Taipei 11696, Taiwan.*

[8]*Department of Medical Research, Shuang Ho Hospital, Taipei Medical University, New Taipei City 23561, Taiwan*

[9]*Department of Pharmacy, Taipei Medical University-Wanfang Hospital, Taipei, Taiwan*

*__Correspondence:__ Wei-Chiao Chang, email: wcc@tmu.edu.tw; Wan-Chen Huang, email: wchuang1@gate.sinica.edu.tw

Abstract

Even though many genetic risk loci for human diseases have been identified and comprehensively cataloged, strategies to guide clinical research by integrating the extensive results of genetic studies and biological resources are still limited. Moreover, integrative analyses that provide novel insights into disease biology are expected to be especially useful for drug discovery. Herein, we use text mining of genetic studies on colorectal cancer (CRC) and assign biological annotations to identified risk genes in order to discover novel drug targets and potential drugs for repurposing. Risk genes for CRC were obtained from PubMed text mining, and for each gene, six functional and bioinformatic annotations were analyzed. The annotations include missense mutations, expression quantitative trait loci (eQTL), molecular pathway analyses, protein-protein interactions (PPIs), genetic overlap with knockout mouse phenotypes, and primary immunodeficiency. We then prioritized biological risk candidate genes according to a scoring system for the six functional annotations. Each functional annotation was assigned one point, and those genes with a score ≥2 were designated "biological CRC risk genes". Using this method, we revealed 82 biological CRC risk genes, which mapped to 128 genes in an expanded PPI network. Further utilizing DrugBank and the Therapeutic Target Database (TTD), we found 21 genes in our list that are targeted by 166 candidate drugs. Based on data from ClinicalTrials.gov, we found our list contains four known target genes with six drugs approved for CRC treatment, as well as three known target genes with nine drugs under preclinical investigation for CRC. Additionally, 12 genes are targeted by 32 drugs approved for other indications, which can possibly be repurposed for CRC treatment. Finally, our analysis from Connectivity Map (CMap) showed that 18 of the 41 drugs under clinical and preclinical investigation have high potential to be useful for CRC.

Keywords: Drug discovery, drug repurposing, data mining, colorectal cancer, bioinformatics

1. Introduction

Colorectal cancer (CRC) is the second most common cause of cancer morbidity and mortality, and it occurs as the fourth most frequent in males and the third most frequent in females globally [1]. Transformation of CRC is the result of a progressive accumulation of genetic and epigenetic alterations. Once the CRC disease is established and diagnosed, patient outcomes are largely dependent on the disease stage, with early-stage disease generally showing better outcomes after treatment, and advanced disease characterized as clinically aggressive [2]. Thus, more effective anticancer drugs are required, as treatment approaches have been limited [2]. Bringing a new drug through clinical trials to the market is time consuming and expensive, often requiring ~15 years and >US$1 billion [3]. Currently, more than 10,000 clinical trials for cancer are registered at www.clinicaltrials.gov, however, only a few drug candidates continue to subsequent phases [4]. In the end, only around 5% of new molecules entering phase I clinical trials will eventually be approved by the U.S. Food and Drug Administration (FDA) for disease treatment [5]. Thus, the approach of drug repurposing, which consists of finding new uses for old drugs as therapeutic agents, has become increasingly popular, including for CRC. Some advantages of drug repurposing are a clear understanding of the drug's safety, pharmacological mechanism and pharmacokinetic profile [6]. One example of clinically successful drug repurposing is the use of aspirin for cardiovascular disease and CRC prevention; aspirin was initially used for pain relief [7, 8]. Additionally, sildenafil was originally used to treat high blood pressure, but was repurposed for erectile dysfunction [9]. Therefore, drug repurposing based on genomic information may be a promising strategy to improve the efficiency of the drug discovery process.

In order to find candidates for drug repurposing, one must identify relationships between diseases and genes, genes and drugs, and diseases and drugs as a primary step [10]. One strategy to identify these relationships is literature searching [10]. For example, candidate genes can be prioritized using keywords to perform text mining on a database. For CRC, an especially useful database may contain large-scale genomic information that may shed light on germline risk alleles in patients with disease [10, 11]. The PubMed database contains more than 24 million citations from MEDLINE and other biomedical literature data sources [12]. In our study we queried the PubMed database to identify candidate risk genes. Then, we utilized six bioinformatics repositories to

identify the genes most likely to have biological impact in CRC etiology, and we also identified relationships between those CRC genes and existing drugs. The current study design was adopted from Okada et al., who predicted drugs for repurposing to rheumatoid arthritis by utilizing risk genes of rheumatoid arthritis and integrating those genetic findings with a complete catalog of approved drugs for rheumatoid arthritis and other diseases [13]. Notably, another genome-wide association study (GWAS) utilized an *in silico* pipeline to study repurposing of drugs to type 2 diabetes; some of the identified drugs had been approved for other diseases [14, 15].

Germline mutations that alter risks of CRC have been previously identified and used to search for target drugs [16, 17]. However, successful translation of this basic research into clinical practice has been limited, and remains as a future challenge. Therefore, in this study, we determined candidate genes associated with CRC and further prioritized the candidate genes using a scoring system based on six *in silico* criteria. As a result, we identified candidate biological risk genes for CRC that provide information to guide the selection of drugs for repurposing to CRC.

2. Materials and Methods

Integrative analyses of genetics-driven genomic drug repurposing for CRC by utilizing multiple databases are shown in **Figure 1**. Single nucleotide polymorphism (SNPs) associated with CRC (CRC risk SNPs) were obtained from PubMed text mining and expanded using HaploReg (v4.1) with a criterion of $r^2 > 0.8$. We identified adjacent SNPs to CRC risk SNPs based on linkage disequilibrium (LD) using HaploReg v4.1, with the criterion of $r^2 > 0.8$; these LD-identified genes are denoted as "CRC-associated genes". Subsequently, genetic data were prioritized based on six functional annotation criteria. Those genes with one functional annotation were awarded one point (score) and those genes with a score ≥2 were classified as "biological CRC genes". The biological CRC genes were utilized in further analyses, including the use of the STRING database [18] to expand the list of candidate genes. We mapped the expanded list of genes according to targets of approved drugs found in the "DrugBank, Therapeutic Target Database. The drug targets were examined to determine their clinical status according to ClinicalTrials.gov. Finally, we used the CMap database to prioritize the most promising candidate drugs for repurposing to CRC treatment.

2.1. Literature search by PubMed text mining

Genes associated with CRC risk were identified through a systematic search of the PubMed database (articles published 2014-2018). The search terms used were: *"colorectal cancer", "risk gene", "susceptibility", "polymorphism", "SNP", and "genotype"*. Searches were performed on December 2, 2018. Inclusion of data was based on the following criteria: 1) all samples were human samples; 2) the clinical phenotype was "CRC"; and 3) the identified SNPs were associated with CRC.

2.2. Biological CRC–associated genes

Expanded SNPs (biological CRC-assiated genes) were identified based on LD using HaploReg v4.1 with a criterion of $r^2 > 0.8$ [19]. LD was used to identify further genes according to the 1000 Genome Project. The LD criterion of $r^2 > 0.8$ was adopted from the study of Okada et al [13].

2.3. Functional annotation of CRC related genes

In order prioritize genes according to their biological function, six annotations were used to construct a scoring system that reflects the most likely candidate genes for use as CRC targets. The first annotation was missense or nonsense mutation according to HaploReg v4.1, which contains annotations of the functional consequences from a database of (db)SNPs [20]. HaploReg v4.1 also links genetic variants to cis-expression quantitative trait loci (cis-eQTL). We examined the cis-eQTL in target tissues of whole blood or colon [20]. In order to understand the connections between mutant genes and phenotypes, we employed WebGestalt 2017 to perform functional enrichment analyses [21]. Further, we used Mammalian Phenotype (MP) Ontology, which contains information on phenotypes of mice and other mammals [22]. To query the mouse phenotype, we converted genes from the human Ensemble ID to mouse Ensemble ID using BioMart [23], and genes with a false discovery rate (*FDR*) of < 0.05 were considered significant results. Protein-protein interaction (PPIs) were identified to understand relationships between diseases and biological protein networks [24]. WebGestalt 2017 was used to perform enrichment analyses and investigate if SNP affected genes were enriched in any specific functional pathway [25]. Specifically, the biological process Gene Ontology (GO) categories in WebGestalt 2017 were interrogated for this step. Significance of a result was set at $FDR < 0.05$ [25]. To determine what kind of molecular pathways were enriched on the CRC-associated gene list and which gene was

involved, enrichment analyses was performed on molecular pathways utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG), an online pathway database of biochemicals from WebGestalt 2017 [25]. Primary immunodeficiency (PID) diseases are innate immune diseases [26], that are often reported to be associated with cancer [27]. PID diseases are known as a genetic disorders associated with an increased incidence of malignancies, such as lymphoproliferative disorders, gastric adenocarcinoma, and CRC [28]. Data were sourced from the International Union of Immunological Societies (IUIS) [26]. A hypergeometric test was used to perform an enrichment analysis on these data; the criterion for significance was p-value < 0.05.

2.4. Prioritization of biological candidate genes

According to these functional annotations, a score was assigned to each gene. One point was awarded for each criteria, as follows: 1) genes with missense or nonsense variants; (2) cis-eQTL genes of risk SNPs; (3) genes prioritized by the Knockout Mouse Phenotype; (4) genes prioritized by PPI network; (5) Molecular Pathway; (6) Primary Immunodeficiency. Each gene was scored according to the number of criteria that were satisfied (scores ranged 0–6 for each gene). Genes with a score of ≥2 were categorized as 'biological CRC risk genes'. We set the threshold of biological score ≥ 2 to find much higher number of the biological CRC genes and candidate of CRC drug targets.

2.5. STRING database

The purpose of the STRING database (http://string-db.org/) is to identify functional interactions related to protein expression by integrating predicted protein-protein association data [18]. A large number of disease-protein networks may provide novel ideas and could be useful for treating various diseases [29]. Genes were filtered from direct PPIs with biological CRC risk genes through the STRING database, version 10.5 (https://string-db.org/cgi/network,). Protein products from the identified biological CRC risk genes, or several genes from a direct PPI network, were explored as potential targets of approved drugs, either for CRC or other indications.

2.6. Drug validation and discovery

The DrugBank database (https://www.drugbank.ca/) and Therapeutic Targets Database (http://bidd.nus.edu.sg/BIDD-Databases/TTD/TTD.asp) were utilized to evaluate potential drug targets. Drug target genes were used to interrogate the databases according to several criteria, such as drugs with pharmacological activities, effectiveness in humans, and annotations of 'approved', 'clinical trial' or 'experimental drugs'. The identified drugs were reviewed on ClinicalTrials.gov (https://clinicaltrials.gov) to identify any clinical investigations for CRC or other diseases.

2.7. Connectivity Map (CMap) Analyses

To prioritize the list of drugs for CRC repurposing, we used the CMap touchstone database platform to rank drugs according to a connectivity score (-100 to 100) [30]. The CMap scores were obtained from the 'Touchstone' tool in the CLUE website (https://clue.io/). The CMap database can be used to perform a profile similiarity analyses, which is based on a collection of genome-wide transcriptional expression profiles from cultured human cancer cells treated with different drug compounds [31]. In this study, drugs were compared with capecitabine as standard treatment for CRC according to the National Comprehensive Cancer Network (NCCN) guideline [32]. The CMap database was a Touchstone dataset containing the signatures from 3000 drugs and genetic loss/gain-of-function for 2000 genes [31]. To identify the similiarity pattern for gene expression of repurposed drugs on CRC-associated cell lines, the drug response was examined for CRC-associated cell lines such as HT29. Drugs with positive connectivity scores induce a similar gene expression profile to the imported query; a negative connectivity score indicates a disimilar profile to that of the imported query.

2.8. Statistical analyses

All analytic workflows were performed using R Studio v3.4.3 and the haploR package (https://www.r-project.org/ and https://cran.r-project.org/web/packages/haploR/index.html). Over-representation analysis (ORA) was used to prioritize the genes in Knockout Mouse Phenotype, PPI network and Molecular Pathway. The *FDR* of <0.05 was determined to indicate the statistical significance. A hypergeometric test was used to perform an enrichment analysis for PID; the criterion for significance was p-value < 0.05.

3. Results

We retrieved SNPs associated with CRC from PubMed text mining. Forty CRC-associated studies were identified that include genotyping studies and GWAS (**Table 1**). In total of 170 unique SNPs were found to associate with CRC risk (**Table S1**). The number of SNPs was expanded based on LD, using HaploReg v4.1 with criterion of $r^2 > 0.8$, and this expansion yielded 210 CRC-associated genes **(Table S2)**.

3.1. Functional annotations of CRC risk SNPs

In order to prioritize genes according to known biological processes, we devised six biological functional annotations based on 210 CRC-associated genes and assigned one point for each functional annotation. The results of six biological functional annotations as shown in **Table S2** and **Figure 2**, we determined that 29 out of 210 genes (10%) were missense or nonsense variants. We then assessed the cis-eQTL effects with HaploReg v4.1 and found 101 genes (33%) risk SNPs have cis-eQTL effects on 210 genes, in either whole blood or colon tissues. Phenotype data were then retrieved from MP Ontology, which contains information on phenotypes of mice and other mammals. WebGestalt 2017 was used to perform an ORA, and 24 genes (8%) were found to overlap with CRC risk genes (*FDR* < 0.05). Gene Ontology (GO) annotations obtained from WebGestalt 2017 were then used to evaluate PPIs. According to the PPIs, 95 genes (31%) that overlapped with other CRC risk genes (*FDR* < 0.05) were found. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to perform an ORA on molecular pathways. A total of 49 CRC-associated genes (16%) in KEGG pathways were successfully identified by the enrichment analysis. We also analyzed PID data from the IUIS and subsequently confirmed overlapping genes with our findings. Five genes (2%) with statistically significant overlap in PID (p <0.05) were identified. After compiling the scores from 0-6**,** we found that genes with a score of 0 and 1 have the same numbers of genes, 64 genes, respectively. Genes with a score of 2 were 32 while those with a score of 3 were 30. There were 16 genes with a score of 4, three genes with a score of 5, and one gene with a score of 6 (**Figure 3**). In total, we obtained eighty-two genes with a score of ≥2, which were categorized as 'biological CRC genes' **(Table 2)**. Our study observed top five of biological CRC genes which were domain-containing protein 2 (NOD2), followed by MutL homolog 1 (MLH1), transforming growth factor beta-1 (TGFB1), tumor protein p53 (TP53) and interleukin 17F (IL17F).

3.2. Expanding the list of biological CRC genes

We used the STRING database of PPIs to expand the number of biological CRC-associated genes. After expansion with the STRING database, 128 genes were included in the list (**Table S3**). These genes comprised the final list of candidate genes, which was used for futher analysis.

3.3. Discovery of CRC drug targets

We obtained 1108 gene pairs from curated PPI networking (**Table S4**) and also found that 21 target genes were existed in DrugBank and the Therapeutic Target Database (**Table S5**). The 166 drugs are either approved, in clinical trials, or experimental drugs for human diseases (**Table S6**). Four target genes with six drugs approved for CRC were identified: bevacizumab, dalteparin, and tegafur uracil (**Figure 4**). Further, correlations of drugs approved for other diseases with biological CRC risk genes were also assessed. We found 12 known targets with 32 drugs for other diseases (under clincal investigation for CRC) that may be repositioned for treatment of CRC (**Figure 5)**. An example is dasatinib, which is approved for treating chronic myeloid leukemia (CML), but has been repurposed for CRC treatment. Another drug, genistein, cytochrome P450 family 1 subfamily a member 1 (CYP1A1) targeting agent, is also a promising drug for treatment of CRC. Genistein may inhibit cancer cell growth by blocking Wnt signaling required for CRC cell growth **(Table S7**). In addition, we also identified five non-steroidal anti-inflammatory drugs (NSAIDs) and four selective serotonin reuptake inhibitors (SSRIs), which are under preclinical *in vitro* or *in vivo* (CRC model) investigations for repurposing to CRC (**Figure 6**). The NSAIDs include balsalazide, tolfenamic acid, adapalene, sulindac, and parecoxib [33-36]. SSRI drugs include amitriptyline, fluoxetine, paroxetine and sertaline [37-40] (**Table S8**). Therefore, our results suggest that this combinational analysis of PubMed text-mining and six functional annotations can efficiently identify potential candidate drugs for CRC drug repurposing.

3.4. Connectivity Map (CMap) analyses prioritize the most promising drugs for CRC

We analyzed 41 drugs under clinical or preclinical investigation using the CMap database to assess their association with known anti-CRC drugs or CRC risk factors. If two drugs have a strongly positive relationship, they may have similar effects in treating CRC. On the contrary, if the relationship is negative, the drugs may have different effects as treatments for CRC. We found that 18 of 41 drugs had positive correlations with capecitabine, the standard treatment for CRC,

according to high CMap scores (**Table 3**). Among the 18 drugs, nine (celecoxib, ibuprofen, naproxen, balsalazide, aspirin, sulindac, rofecoxib, diclofenac and parecoxib) target cyclooxygenase (COX)-1 and COX-2, three (sertraline, fluoxetine and amitriptyline) target SLC6A4, two (rucaparib and olaparib) target poly (ADP-ribose) polymerase (PARP), and one (dasatinib) targets cyclin-D2 (CCND2). As depicted in **Table 3**, the top five drugs that exhibited strong relationships with capecitabine were celecoxib, rucaparib, ibuprofen, naproxen and balsalazide, with respective CMap scores of 99.09, 94.14, 92.25, 86.40, and 81.55. This finding suggests the biological plausibility of repurposing cancer drugs from non-cancer drugs or other drugs to treatment of CRC.

4. Discussion

In the current study, we mined PubMed text to extract potential candidate genes that may guide drug repurposing to CRC. Six functional annotations were applied to build a scoring system and to prioritize CRC risk genes likely to be useful drug targets. We found 128 genes from the expanded PPI network, including four known targets of six approved CRC drugs, 12 known targets for 32 drugs approved for other indications (with potential to be repurposed for CRC), and three known target genes under preclinical investigation for CRC that may be modulated by nine drugs. Finally by using the CMap database, we observed 18 of the 41 drugs under clinical and preclinical investigations that might have high potential to be repurposed for CRC. We found six drugs, including genistein, marimastat, bevacizumab, dalteparin, aflibercept, and tegafur-uracil, that are currently in clinical trials for CRC. Genistein [NCT01985763] and Aflibercept [NCT03264274] are currently in a phase 2 clinical trial for treatment of metastatic CRC. Xeloda or UFT (tegafur-uracil) with folinic acid is in a phase 3 investigation for metastatic CRC [NCT00905047]. Bevacizumab is in a phase 2 clinical trial for advanced CRC in combination with irinotecan, lederfolin, and 5FU (FOLFIRI) [NCT01853813]. Dalteparin was assessed by a double-blinded phase 3 randomized control trial for treating advanced cancers, including breast, lung and colorectal [NCT00003674].

Although adjuvant therapy is available for patients with high-risk stage II and stage III CRC, around 40-50% of patients will experience recurrence with the development of metastatic disease [41]. For advanced CRC, targeting the vascular endothelial growth factor (VEGF) pathway has a

pivotal role in increasing overall survival time in patients receiving chemotherapy [42]. Herein, VEGF is one of the drug targets identified by our analysis; VEGF can be targeted by bevacizumab, dalteparin, and aflibercept. Two of these drugs (bevacizumab and aflibercept) have been approved for use in metastatic CRC [43]. Indeed, most drugs that we found have indications for other cancer types. The 32 drugs in this category are depicted in **Figure 5** (e.g., ovarian cancer [rucaparib, niraparib, olaparib], non-small cell lung cancer [pemetrexed], prostate cancer [abiraterone], osteosarcoma [mifamurtide], CML [dasatinib], multiple myeloma [pomalidomide]). Another set of drugs identified in our study are currently used for autoimmune disease (e.g., ankylosing spondylitis [etanercept, adalimumab, infliximab, golimumab, secuximab]). Surprisingly, one drug we identified as potentially useful to for CRC is vitamin D. Vitamin D deficiency may affect various human cancers, and the Vitamin D receptor (VDR), which plays a role in regulating active vitamin D metabolites, is highly expressed in the small intestine and colon [44]. Vitamin D has been shown to protect from CRC tumorigenesis by binding long-chain fatty acids and bile acids in the small intestine and protecting colonic epithelial cells from mutagens [45]. Notably, Vitamin D-related drugs are currently being evaluated in CRC: cholecalciferol is in a phase 2 trial [NCT01074216]; cholecalciferol [NCT02172651] and paricalcitol [NCT01197664] are in phase 1 trials. Furthermore, we identified calcitriol as a promising drug for CRC repurposing based on the CMap analysis (**Table 3**). However, molecular evidence for the role of vitamin D in CRC is limited, so future functional studies are still needed to investigate the mechanisms by which vitamin D affects CRC.

Another category of drugs identified in our study is NSAIDs, which generally hold indications as analgesic, antipyretic and anti-inflammatory agents [46]. From the results of a nationwide population-based case–control study, Kuo et al., indicated that NSAIDs use was associated with lower incidence of colorectal cancer [47]. According to the CMap analysis we performed, nine NSAIDs (celecoxib, ibuprofen, naproxen, balsalazide, aspirin, sulindac, rofecoxib, diclofenac and parecoxib) were among the most promising for CRC repurposing. Furthermore, NSAIDs were also reported to reduce the growth of cancer cells through inhibition of two enzymes, COX-1 and COX-2 [48]. Aspirin is commonly used in cardiovascular disease treatment and to prevent myocardial infarction. Previous clinical studies indicated that post-diagnosis use of aspirin is correlated with favorable survival in CRC patients [49, 50]. Ibuprofen was also reported to effectively inhibit the

growth of prostate cancer [51]. In an *in vitro* study on adenocarcinoma gastric cells, the antitumor activity of ibuprofen was attributed to the induction of apoptosis and the reduction of cell proliferation, which attenuated angiogenesis [52]. Naproxen is a propionic-acid derivative and non-selective COX-inhibitor that was shown to inhibit cell proliferation, trigger apoptosis and suppress metastasis. *In vitro* and *in vivo* studies revealed that naproxen also had antineoplastic properties in leukemia, breast, colon, bladder, and osteosarcoma cell lines [53]. Furthermore, according to an *in vivo* study on a colon cancer animal model, the combination of naproxen and atorvastation significantly inhibited growth of colonic adenocarcinomas [54]. Diclofenac is a derivative of acetic acid that has a moderate selectivity for COX-2 inhibition. Diclofenac was reported to have antineoplastic effects in various tumor types, including fibrosarcomas, hepatomas, and colon, ovarian and pancreatic cancer [55]. Diclofenac also was reported to have anti-proliferative effects in human colon cancer cell lines [56]. Interestingly, other NSAIDs that we identified are celecoxib and rofecoxib as drugs with the potential to be repurposed for CRC. A previous study of celecoxib in an animal model showed that it is effective at reducing colorectal adenomas in animal models and in patients with familial adenomatous polyposis (FAP) [57]. Celecoxib and rofecoxib are two COX-2 inhibitors approved by the U.S. FDA for FAP patients, and these drugs are also being studied in patients with sporadic adenomas and other cancer types [57]. A study of celecoxib as an adjuvant chemotherapy is ongoing in patients with metastatic CRC (phase 4) [NCT03645187]. Interestingly, Yang et al., reported that it may be effective for treating ovarian cancer [12], and Kim et al. suggested that the combination of celecoxib and paclitaxel could be an effective treatment for ovarian cancer [58]. Moreover, some NSAIDs are being evaluated in preclinical models for repurposing to CRC (**Figure 6**); such drugs include balsalazide, tolfenamic acid, adapalene and sulindac, which inhibited the proliferation and induced apoptosis in human colon cancer cells [33-36].

Additionally, CMap analysis also showed that the SSRIs drugs, sertraline, fluoxetine and amitriptyline, may be promising candidates for repurposing to CRC. SSRI drugs have been widely used for the treatment of depression and anxiety [59]. Recently, several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [37-40]. Stopper et al., reported that fluoxetine may reduce tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage

in tumor metabolism [38]. A following investigation also supported the anticancer activity of sertraline in the two human colon cancer cell lines (HT29 and LS1034) [40]. Currently, biomarkers for CRC are still limited to kirsten rat sarcoma (KRAS), neuroblastoma RAS (NRAS) and B-Raf proto-oncogene (BRAF), with the upcoming standard of care including phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), human epidermal growth factor receptor 2 **(**HER-2) and microsatellite instability (MSI) [43]. PIK3CA is one of the therapeutic targets we found in this study and may be targeted by copanlisib, which has an original indication for follicular lymphoma. Copanlisib, a phosphatidylinositol 3-kinase (PI3K) inhibitor, combined with nivolumab, an anti-PD-1 agent, is under evaluation in patients with mismatch-repair proficient (MSS) colorectal cancer [NCT03711058]. Furthermore, the safety and efficacy of copanlisib in combination with nivolumab are being assessed in an on-going phase 1b/2 study [NCT03735628].

Our study provides new insights into how oncogenomic research may be implemented clinically, based on a genomics approach to guide selection of CRC pharmacotherapy agents. The advantage of drug repurposing is that the safety data in human patients for identified drugs are already available [60]. This *in silico* approach might be useful to narrow down candidate drugs for CRC disease, resulting in cost and time savings. ~~However, further functional studies will be needed to examine the drugs revealed in our analyses.~~ In particular, we would like to emphasize that CRC genetic-driven genomic drug repurposing is a new model of leveraging CRC genomic information, not only to identify CRC disease risk but also to provide novel biological insight and contribute to CRC drug repurposing. However, our approaches have some limitations. The target genes that we identified are not all in pharmacological activities, therefore these might potentially miss the target of the drugs (undruggable), our analysis showed, only 166 drugs (14.98%) of the 1108 protein coding genes are druggable (genetic driven druggable). The comprehensive analysis on the druggability of genes performed by Finan et al. also revealed that only 4479 (22%) among 20,300 protein coding genes are druggable [61]. Moreover, functional studies of the biology of these risk genes and the genes targeted by these drugs still require further investigation to ascertain the role of drug target genes.

5. Conclusion

By utilizing genomic data to map disease-gene-protein-drug relationships, we found 82 biological CRC risk genes and 128 genes from an expanded PPI network. Among the identified targets, four genes are the targets of six drugs already approved for CRC treatment, three genes are targeted by nine drugs under preclinical assessment for CRC, and 12 target genes overlap with 32 drugs approved for other indications; these drugs with other indications might be good candidates for repurposing to CRC treatment. Moreover, our CMap analysis indicates that 18 of the 41 drugs under clinical and preclinical investigation demonstrate potential to be repurposed for CRC treatment. Thus, our study demonstrates the feasibility of utilizing genomic data and genetic mapping as a potential method for drug discovery.

Author Contributions

L.M.I., H.S.C.W., W.H.C and W.C.C conceived and designed the study. L.M.I performed the computational analysis. L.M.I wrote the manuscript. W.C.C provided the funding. H.S.C.W., W.H.C., W.C.H., W.A., E.M., and W.C.C revised the manuscript. W.C.H and W.C.C supervised and coordinated this study. All authors have read and approved of the manuscript and have made significant contributions to this study

Declaration of Competing Interest

The authors disclose no conflict of interest

Data Availability

The dataset used and/or analysed in the study are obtained from ClinicalTrials.gov (https://clinicaltrials.gov/); CMap 'Touchstone' tool in the CLUE website (https://clue.io/); Drug bank databases (https://www.drugbank.ca/); HaploReg (v4.1) (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php); STRING database (http://string-db.org/); Therapeutic Targets Database (http://bidd.nus.edu.sg/BIDD-Databases/TTD/TTD.asp); WebGestalt 2017 (http://www.webgestalt.org/2017/option.php); PubMed database (https://pubmed.ncbi.nlm.nih.gov/).

Acknowledgements

This project was supported by grants from the Ministry of Science and Technology, Taiwan (MOST108-2314-B-038-027), The National Health Research Institutes, Taiwan (08A1-MGGP08-037), and Taipei Medical University (106-5807-001-400).

Abbreviations

BRAF	: B-Raf proto-oncogene
CCND2	: Cyclin-D2
CRC	: Colorectal cancer
CMap	: Connectivity map
COX-1	: Cyclooxygenase-1
COX-2	: Cyclooxygenase-2
CML	: Chronic myeloid leukemia
Cis e-QTL	: Cis expression~quantitative trait loci
CYP1A1	: Cytochrome P450 Family 1 Subfamily A Member 1
FDA	: Food and Drug Administration
FAP	: Familial adenomatous polyposis
GWAS	: Genome-wide association study
GO	: Gene ontology
HER-2	: Human epidermal growth factor receptor 2
IUIS	: International Union of Immunological Societies
IL17F	: Interleukin 17F
KEGG	: Kyoto Encyclopedia of Genes and Genomes
KRAS	: Kirsten rat sarcoma
LD	: Linkage disequilibrium
MLH1	: MutL homolog 1
MSI	: Microsatellite instability
NHGRI	: National Human Genome Research Institute
NOD2	: Domain-containing protein 2
NSAIDs	: Nonsteroidal anti-inflammatory drugs
NRAS	: Neuroblastoma RAS

PARP	: Poly (ADP-ribose) polymerase
PMID	: PubMed IDentifier
PPI	: Protein-protein interaction
PID	: Primary immunodeficiency
PI3K	: Phosphatidylinositol 3-kinase
PIK3CA	: Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
SNPs	: Single nucleotide polymorphism
SSRIs	: Selective serotonin-reuptake inhibitors
TGFB1	: Transforming growth factor beta-1
TP53	: Tumor protein p53
VEGF	: Vascular endothelial growth factor
VDR	: Vitamin D receptor
WES	: Whole exome sequencing

Supporting Information

Additional file 1: Table S1. Unique colorectal cancer (CRC)-associated SNPs. (DOCX)

Additional file 2: Table S2. Prioritization of CRC-associated genes based on biological annotations. (DOCX)

Additional file 3: Table S3. Colorectal cancer (CRC)-related genes after expansion with the STRING database. (DOCX)

Additional file 4: Table S4. Genes pairs from STRING database. (XLSX)

Additional file 5: Table S5. List of 21 target genes for CRC. (XLSX)

Additional file 6: Table S6. List of 166 drugs for 21 known targets. (XLSX)

Additional file 7: Table S7. Drugs under clinical investigation for colorectal cancer (CRC). (DOCX)

Additional file 8: Table S8. Drugs under preclinical investigation for colorectal cancer (CRC). (DOCX)

References

[1] R. Siegel, J. Ma, Z. Zou, *et al.*, Cancer statistics, 2014, CA: a cancer journal for clinicians 64(1) (2014) 9-29.

[2] H.J. Freeman, Early stage colon cancer, World J Gastroenterol 19(46) (2013) 8468-73.

[3] M. Sirota, J.T. Dudley, J. Kim, *et al.*, Discovery and preclinical validation of drug indications using compendia of public gene expression data, Science translational medicine 3(96) (2011) 96ra77.

[4] M. Hay, D.W. Thomas, J.L. Craighead, *et al.*, Clinical development success rates for investigational drugs, Nature biotechnology 32(1) (2014) 40-51.

[5] S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, *et al.*, How to improve R&D productivity: the pharmaceutical industry's grand challenge, Nature reviews. Drug discovery 9(3) (2010) 203-14.

[6] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, Nature reviews. Drug discovery 3(8) (2004) 673-83.

[7] A.M. Algra, P.M. Rothwell, Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials, The Lancet. Oncology 13(5) (2012) 518-27.

[8] T. Wolff, T. Miller, S. Ko, Aspirin for the primary prevention of cardiovascular events: an update of the evidence for the U.S. Preventive Services Task Force, Annals of internal medicine 150(6) (2009) 405-10.

[9] H.A. Ghofrani, I.H. Osterloh, F. Grimminger, Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond, Nature reviews. Drug discovery 5(8) (2006) 689-702.

[10] Y. Luo, G. Riedlinger, P. Szolovits, Text mining in cancer gene and pathway prioritization, Cancer informatics 13(Suppl 1) (2014) 69-79.

[11] A. Singhal, M. Simmons, Z. Lu, Text Mining Genotype-Phenotype Relationships from Biomedical Literature for Database Curation and Precision Medicine, PLoS computational biology 12(11) (2016) e1005017.

[12] H.T. Yang, J.H. Ju, Y.T. Wong, *et al.*, Literature-based discovery of new candidates for drug repurposing, Briefings in bioinformatics 18(3) (2017) 488-497.

[13] Y. Okada, D. Wu, G. Trynka, *et al.*, Genetics of rheumatoid arthritis contributes to biology and drug discovery, Nature 506(7488) (2014) 376-81.

[14] M. Imamura, A. Takahashi, T. Yamauchi, *et al.*, Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes, Nature communications 7 (2016) 10531.

[15] J. Zhang, K. Jiang, L. Lv, *et al.*, Use of genome-wide association studies for cancer research and drug repositioning, PloS one 10(3) (2015) e0116477.

[16] E.M. Stoffel, E. Koeppe, J. Everett, *et al.*, Germline Genetic Features of Young Individuals With Colorectal Cancer, Gastroenterology 154(4) (2018) 897-905.e1.

[17] S.A. Smith, T. French, S.J. Hollingsworth, The impact of germline mutations on targeted therapy, The Journal of Pathology 232(2) (2014) 230-243.

[18] D. Szklarczyk, J.H. Morris, H. Cook, *et al.*, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res 45(D1) (2017) D362-d368.

[19] I.Y. Zhbannikov, K. Arbeev, S. Ukraintseva, *et al.*, haploR: an R package for querying web-based annotation tools, F1000Research 6 (2017) 97.

[20] L.D. Ward, M. Kellis, HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease, Nucleic Acids Res 44(D1) (2016) D877-81.
[21] J. Wang, S. Vasaikar, Z. Shi, *et al.*, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit, Nucleic Acids Res 45(W1) (2017) W130-w137.
[22] C.L. Smith, J.T. Eppig, The mammalian phenotype ontology: enabling robust annotation and comparative analysis, Wiley interdisciplinary reviews. Systems biology and medicine 1(3) (2009) 390-399.
[23] S. Durinck, Y. Moreau, A. Kasprzyk, *et al.*, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis, Bioinformatics (Oxford, England) 21(16) (2005) 3439-40.
[24] J. De Las Rivas, C. Fontanillo, Protein-protein interactions essentials: key concepts to building and analyzing interactome networks, PLoS computational biology 6(6) (2010) e1000807.
[25] M. Ashburner, C.A. Ball, J.A. Blake, *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nature genetics 25(1) (2000) 25-9.
[26] N. Parvaneh, J.L. Casanova, L.D. Notarangelo, *et al.*, Primary immunodeficiencies: a rapidly evolving story, The Journal of allergy and clinical immunology 131(2) (2013) 314-23.
[27] E. Mortaz, P. Tabarsi, D. Mansouri, *et al.*, Cancers Related to Immunodeficiencies: Update and Perspectives, Front Immunol 7 (2016) 365.
[28] K. Salavoura, A. Kolialexi, G. Tsangaris, *et al.*, Development of cancer in patients with primary immunodeficiencies, Anticancer research 28(2b) (2008) 1263-9.
[29] H. Ruffner, A. Bauer, T. Bouwmeester, Human protein-protein interaction networks and the value for drug discovery, Drug discovery today 12(17-18) (2007) 709-16.
[30] A. Subramanian, R. Narayan, S.M. Corsello, *et al.*, A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles, Cell 171(6) (2017) 1437-1452.e17.
[31] J. Lamb, E.D. Crawford, D. Peck, *et al.*, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease, Science 313(5795) (2006) 1929-35.
[32] N.C.C. Network, NCCN Guideline for Colon Cancer, (2018) 26.
[33] H.-Y. Kim, S.-L. Kim, Y.-R. Park, *et al.*, Balsalazide Potentiates Parthenolide-Mediated Inhibition of Nuclear Factor-κB Signaling in HCT116 Human Colorectal Cancer Cells, Intestinal Research 13(3) (2015) 233-241.
[34] S.-H. Lee, J.H. Bahn, C.K. Choi, *et al.*, ESE-1/EGR-1 pathway plays a role in tolfenamic acid-induced apoptosis in colorectal cancer cells, Molecular cancer therapeutics 7(12) (2008) 3739-3750.
[35] H.N. Tinsley, B.D. Gary, J. Thaiparambil, *et al.*, Colon tumor cell growth-inhibitory activity of sulindac sulfide and other nonsteroidal anti-inflammatory drugs is associated with phosphodiesterase 5 inhibition, Cancer prevention research (Philadelphia, Pa.) 3(10) (2010) 1303-1313.
[36] W. Xiong, W.H. Li, Y.X. Jiang, *et al.*, Parecoxib: an enhancer of radiation therapy for colorectal cancer, Asian Pac J Cancer Prev 16(2) (2015) 627-33.
[37] H. Arimochi, K. Morita, Characterization of cytotoxic actions of tricyclic antidepressants on human HT29 colon carcinoma cells, European journal of pharmacology 541(1-2) (2006) 17-23.
[38] H. Stopper, S.B. Garcia, A.M. Waaga-Gasser, *et al.*, Antidepressant fluoxetine and its potential against colon tumors, World J Gastrointest Oncol 6(1) (2014) 11-21.

[39] W.J. Jang, S.K. Jung, T.T.L. Vo, *et al.*, Anticancer activity of paroxetine in human colon cancer cells: Involvement of MET and ERBB3, Journal of cellular and molecular medicine 23(2) (2019) 1106-1115.
[40] I. Gil-Ad, A. Zolokov, L. Lomnitski, *et al.*, Evaluation of the potential anti-cancer activity of the antidepressant sertraline in human colon cancer cell lines and in colorectal cancer-xenografted mice, International journal of oncology 33(2) (2008) 277-86.
[41] S. Gill, A.W. Blackstock, R.M. Goldberg, Colorectal cancer, Mayo Clinic proceedings 82(1) (2007) 114-29.
[42] B. Gustavsson, G. Carlsson, D. Machover, *et al.*, A review of the evolution of systemic chemotherapy in the management of colorectal cancer, Clinical colorectal cancer 14(1) (2015) 1-10.
[43] P.J. Cuyle, H. Prenen, Current and future biomarkers in the treatment of colorectal cancer, Acta clinica Belgica 72(2) (2017) 103-115.
[44] J. Sun, The Role of Vitamin D and Vitamin D Receptors in Colon Cancer, Clin Transl Gastroenterol 8(6) (2017) e103-e103.
[45] S.A. Lamprecht, M. Lipkin, Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis, Annals of the New York Academy of Sciences 952 (2001) 73-87.
[46] J.K. Buer, Origins and impact of the term 'NSAID', Inflammopharmacology 22(5) (2014) 263-7.
[47] C.N. Kuo, J.J. Pan, Y.W. Huang, *et al.*, Association between Nonsteroidal Anti-Inflammatory Drugs and Colorectal Cancer: A Population-Based Case-Control Study, Cancer epidemiology, biomarkers & prevention 27(7) (2018) 737-745.
[48] J.Y. Wang, J. Sun, M.Y. Huang, *et al.*, STIM1 overexpression promotes colorectal cancer progression, cell motility and COX-2 expression, Oncogene 34(33) (2015) 4358-67.
[49] M.S. Reimers, E. Bastiaannet, M.P. van Herk-Sukel, *et al.*, Aspirin use after diagnosis improves survival in older adults with colon cancer: a retrospective cohort study, Journal of the American Geriatrics Society 60(12) (2012) 2232-6.
[50] C. McCowan, A.J. Munro, P.T. Donnan, *et al.*, Use of aspirin post-diagnosis in a cohort of patients with colorectal cancer and its association with all-cause and colorectal cancer specific mortality, European journal of cancer 49(5) (2013) 1049-57.
[51] M.H. Kim, J. Chung, Synergistic cell death by EGCG and ibuprofen in DU-145 prostate cancer cell line, Anticancer research 27(6b) (2007) 3947-56.
[52] H. Akrami, S. Aminzadeh, H. Fallahi, Inhibitory effect of ibuprofen on tumor survival and angiogenesis in gastric cancer cell, Tumour biology 36(5) (2015) 3237-43.
[53] M.S. Kim, J.E. Kim, D.Y. Lim, *et al.*, Naproxen induces cell-cycle arrest and apoptosis in human urinary bladder cancer cell lines and chemically induced cancers by targeting PI3K, Cancer prevention research (Philadelphia, Pa.) 7(2) (2014) 236-45.
[54] N. Suh, B.S. Reddy, A. DeCastro, *et al.*, Combination of atorvastatin with sulindac or naproxen profoundly inhibits colonic adenocarcinomas by suppressing the p65/beta-catenin/cyclin D1 signaling pathway in rats, Cancer prevention research (Philadelphia, Pa.) 4(11) (2011) 1895-902.
[55] L. Sleire, H.E. Forde, I.A. Netland, *et al.*, Drug repurposing in cancer, Pharmacological research 124 (2017) 74-91.
[56] A.V. Smirnova, L.B. Lazebnik, I.E. Trubitsina, *et al.*, [Antiproliferative activity of diclofenac at tumor cell cultures], Eksperimental'naia i klinicheskaia gastroenterologiia = Experimental & clinical gastroenterology (5) (2012) 66-9.

[57] A.J. Chu, T.H. Chou, B.D. Chen, Prevention of colorectal cancer using COX-2 inhibitors: basic science and clinical applications, Frontiers in bioscience 9 (2004) 2697-713.
[58] H.J. Kim, G.W. Yim, E.J. Nam, *et al.*, Synergistic Effect of COX-2 Inhibitor on Paclitaxel-Induced Apoptosis in the Human Ovarian Cancer Cell Line OVCAR-3, Cancer research and treatment 46(1) (2014) 81-92.
[59] J. Ramsberg, C. Asseburg, M. Henriksson, Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model, PloS one 7(8) (2012) e42003-e42003.
[60] M.R. Hurle, L. Yang, Q. Xie, *et al.*, Computational drug repositioning: from data to therapeutics, Clinical pharmacology and therapeutics 93(4) (2013) 335-41.
[61] C. Finan, A. Gaulton, F.A. Kruger, *et al.*, The druggable genome and support for target identification and validation in drug development, Science Translational Medicine 9(383) (2017) eaag1166.

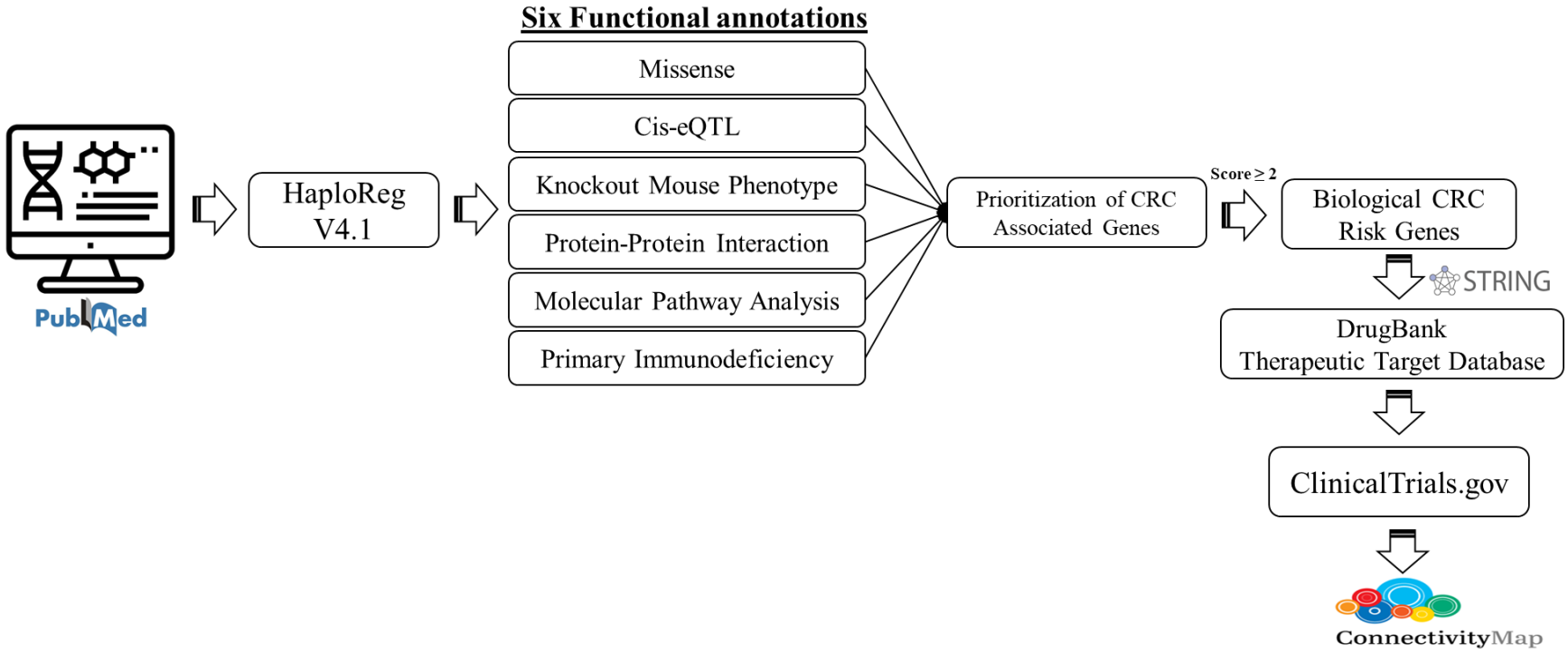


Figure 1. Flow chart for the colorectal cancer (CRC) drug repurposing study. CRC-associated single nucleotide polymorphism (SNPs) were identified by searching PubMed literature. The list of CRC-associated SNPs was expanded with HaploReg v4.1 to find additional CRC-associated genes. Six functional annotations criteria were used to generate a biological prioritization score: (1) CRC risk missense variant; (2) Cis-eQTL; (3) knockout mouse phenotype; (4) Protein-Protein Interaction; (5) molecular pathway analysis; and (6) primary immunodeficiency. Each gene was scored according to the number of criteria that were satisfied (scores ranged 0–6 for each gene). Genes with a score of ≥2 were categorized as 'biological CRC risk genes'. The biological CRC genes were utilized in further analyses, including the use of the STRING database to expand the list of candidate genes. The extent of overlap with target genes for approved CRC drugs was also assessed using the DrugBank database and Therapeutic Target Database (TTD). Drugs in clinical trials were identified using ClinicalTrials.gov. Finally, the connectivity map (CMap) database was used to prioritize the most promising drugs for CRC treatment.

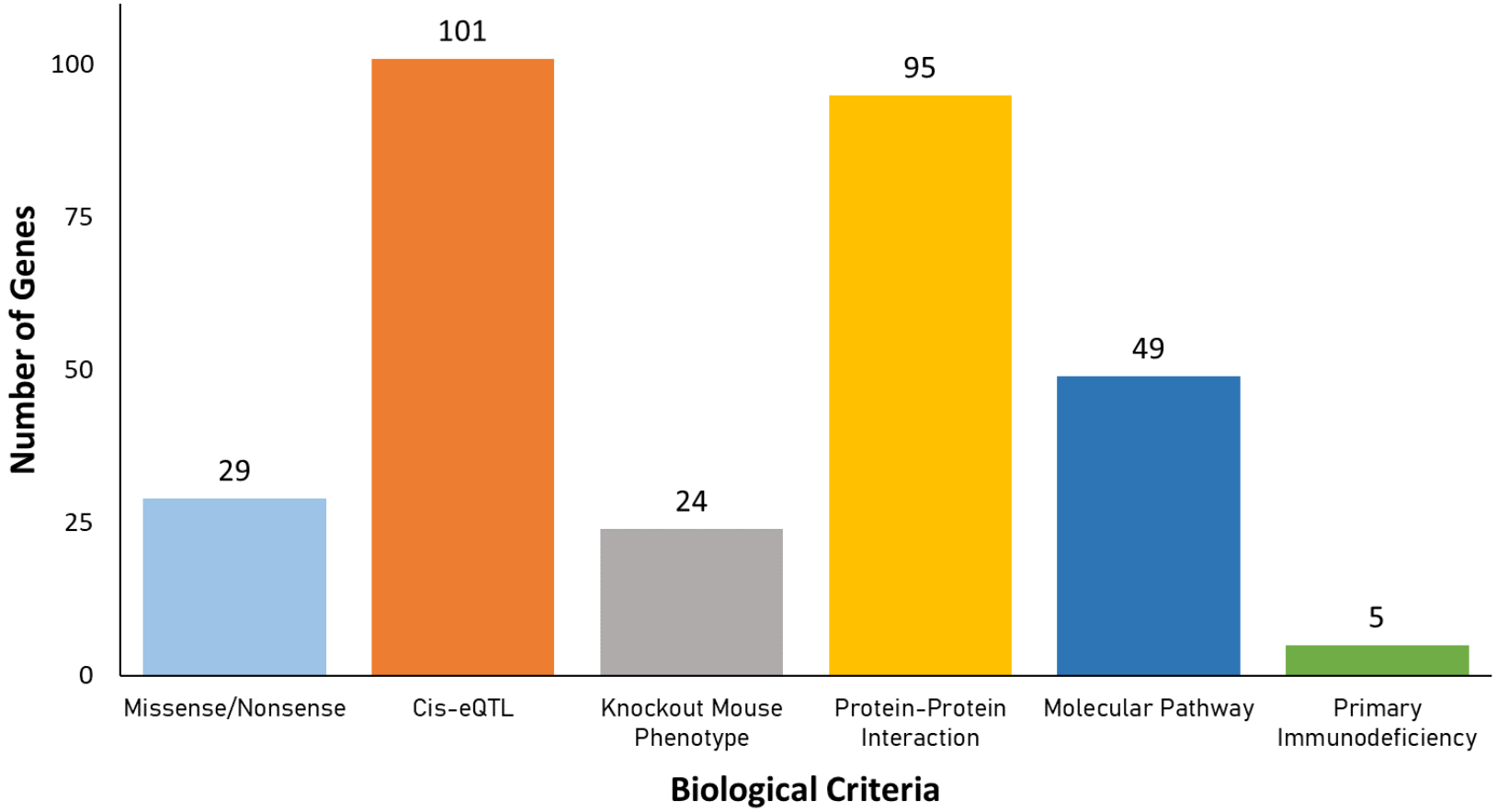


Figure 2. The number of genes satisfying each criterion. The figure, from left- to the right-side, showed 29 genes which were missense or nonsense variants (blue), 101 genes have Cis-expression quantitative trait loci (Cis-eQTL) effects on whole blood or colon (orange), 24 genes with knockout mouse phenotype (grey), 95 genes with protein-protein interaction (yellow), 49 CRC-associated genes in molecular pathways (light blue) and five genes overlap in primary immunodeficiency (green).

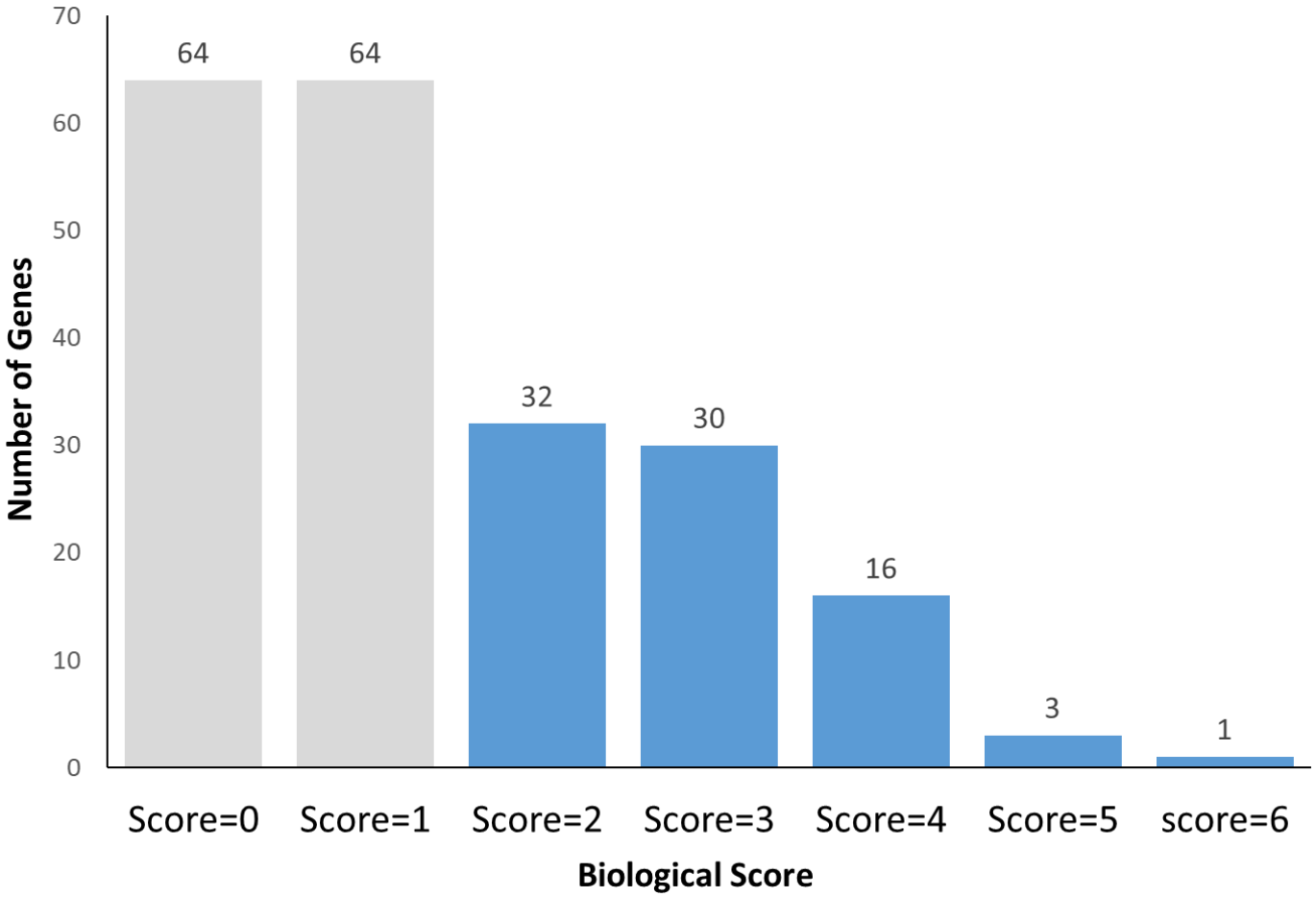


Figure 3. Distribution of colorectal cancer (CRC)-related gene scores. The figure, from left- to the right-side, showed number of genes with score 0~6. Each gene was assigned one point for each functional annotation. Genes with a score of 0 and 1 have the same numbers, namely 64 genes. Genes with a score of 2 were 32 while those with a score of 3 were 30. There were 16 genes with a score of 4, 3 genes with a score of 5, and 1 gene with a score of 6. After compiling the scores, eighty-two genes with a score of ≥2, which were categorized as 'biological CRC genes'.

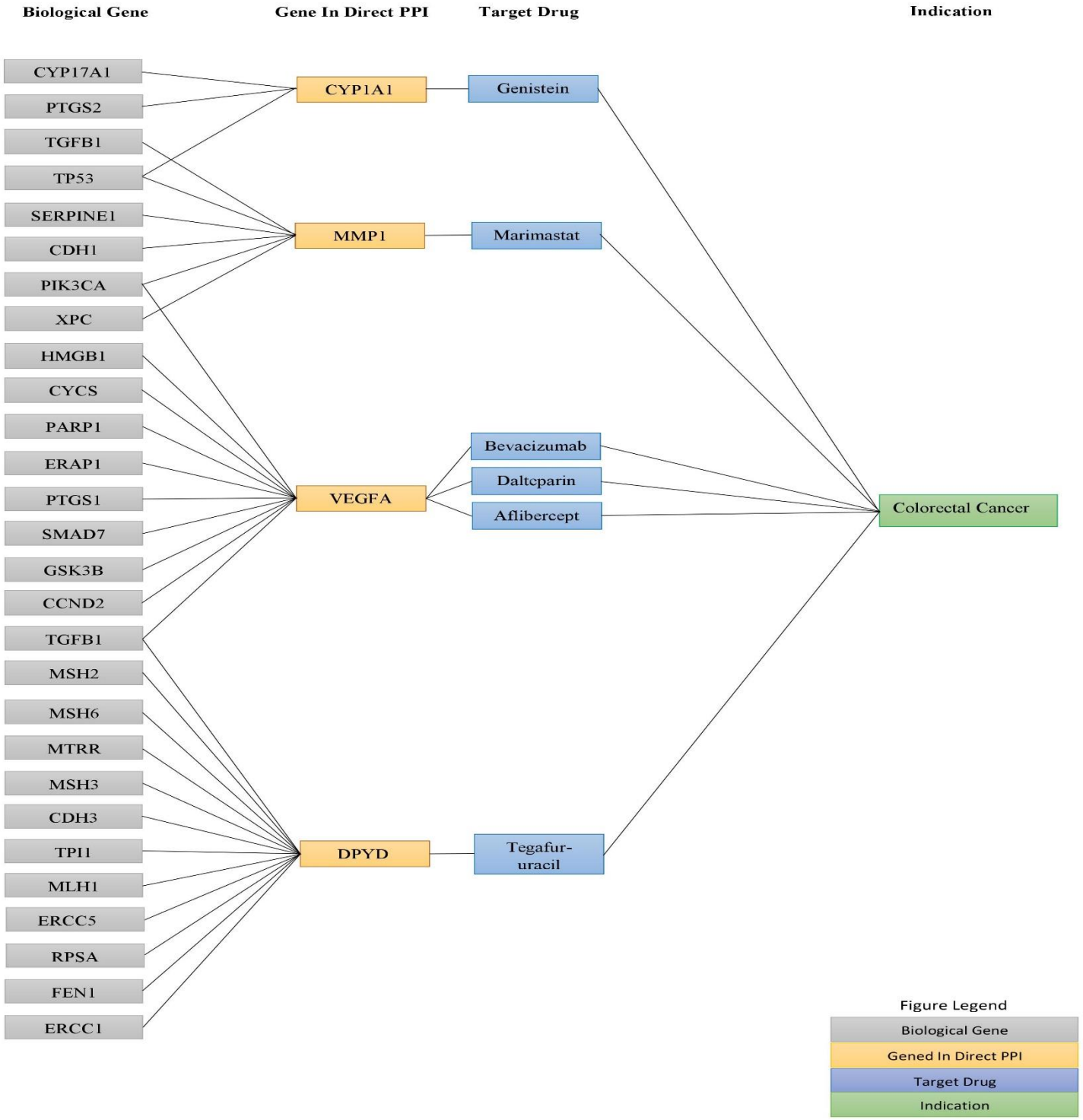


Figure 4. Connections between biological colorectal cancer (CRC) candidate genes and other genes, according to direct protein-protein ineteractions (PPIs), as well as drugs available for CRC target the candidate gene. Biological genes (grey); genes in PPIs (light yellow); target drugs (blue); indication (green).

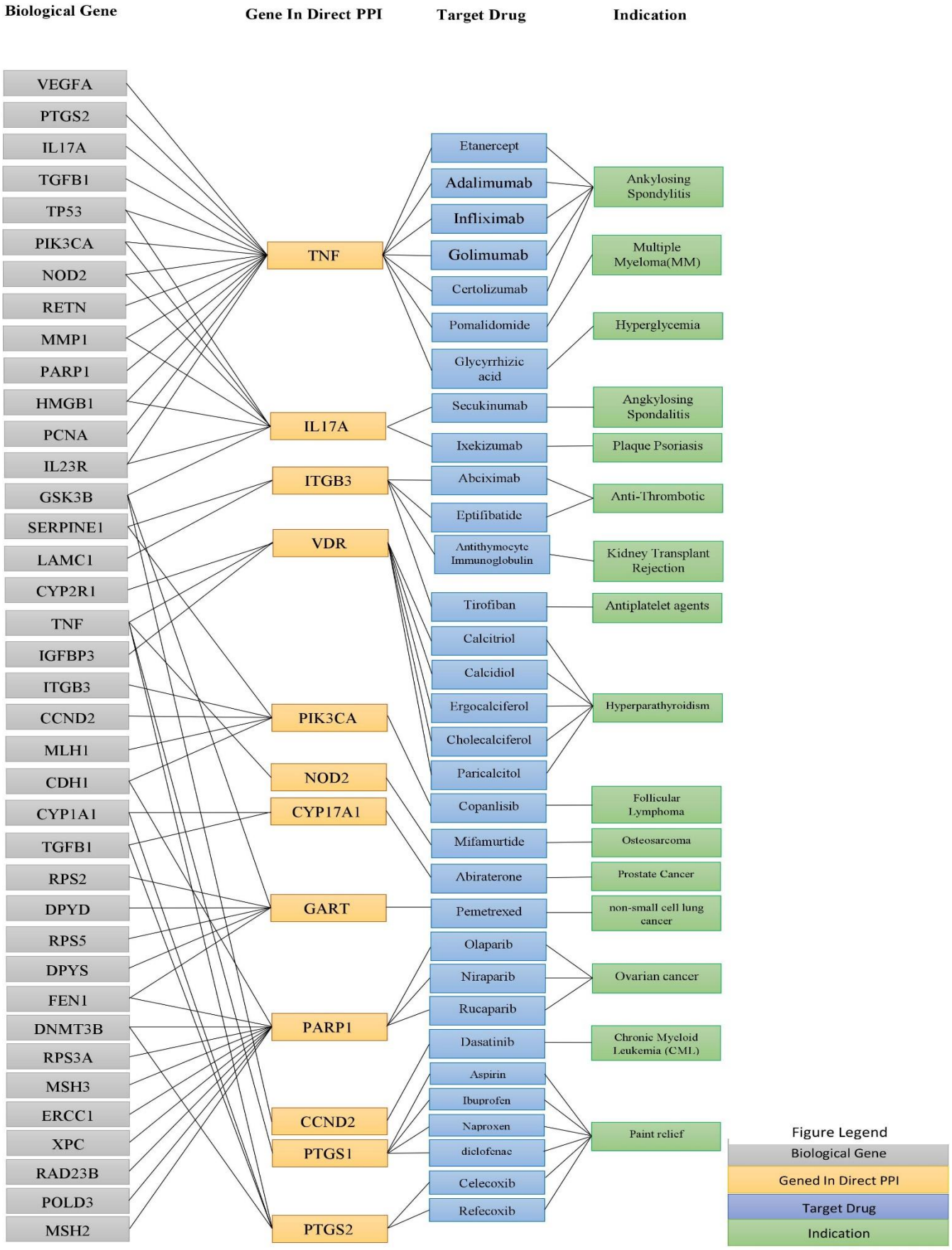


Figure 5. Connections between CRC candidate genes and drugs with other indications (under clinical investigation for CRC). Biological genes (grey); genes in PPI (light yellow); target drugs(blue); indication (green).

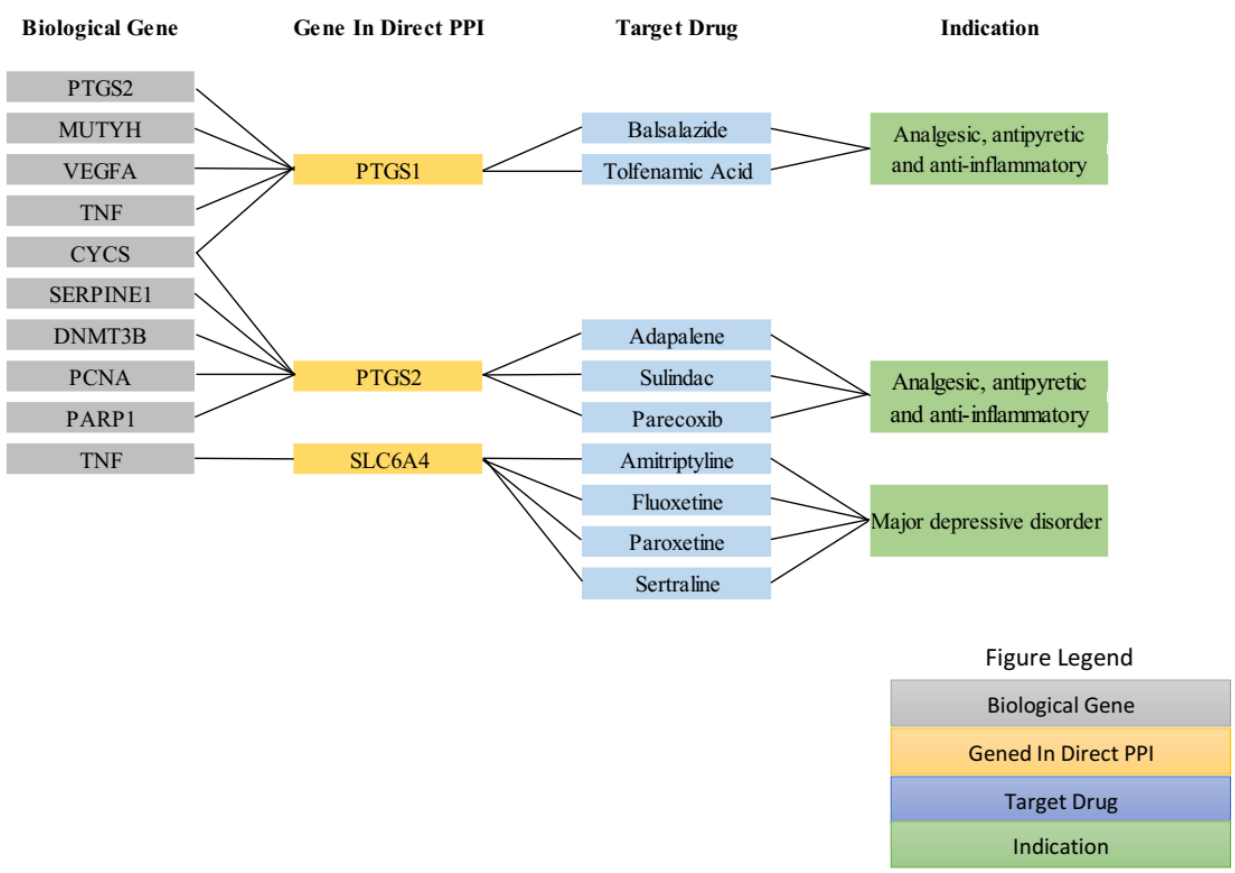


Figure 6. Connections between colorectal cancer (CRC) candidate genes and drugs under preclinical investigation for CRC. Biological genes (grey); genes in PPI (light yellow); target drugs (blue); and indication (green).

Table 1. Profiles of recruited colorectal cancer (CRC) studies

No	PMID	Type of the Study	Population	Disease
1	30018674	WES	Chinese	CRC
2	28915899	GWAS	Italian	CRC
3	24737748	GWAS	European	CRC
4	25811207	Genotype	Saudi Arabian	CRC
5	26965516	GWAS	East Asian, Chinese, Japanese Korean	CRC
6	24836286	GWAS	East Asian	CRC
7.	30233234	Genotyping	Saudi Arabian	CRC
8.	26515597	GWAS	Chinese	CRC
9	25105248	GWAS	Japanese	CRC
10	27698911	Genotyping	Chinese	CRC
11	27145994	GWAS	Chinese	CRC
12	28218435	GWAS	Chinese	CRC
13	29267898	ChIP-Seq, DNase-Seq	Chinese	CRC
14	24448986	GWAS	East Asian	CRC
15	23875689	Genotyping	Korean	CRC
16	23946381	GWAS	European-Asian	CRC
17	25990418	GWAS	European	CRC
18	29428571	Genotyping	Chinese	CRC
19	24968322	Genotyping	Taiwanese	CRC
20	29689450	Genotyping	Tunisians	CRC
21	29953646	Genotyping	Iran	CRC
22	29372689	Genotyping	European	CRC
23	29419695	Genotyping	Chinese	CRC
24	27146020	GWAS	NHGRI	CRC
25	29766219	Genotyping	Chinese	CRC
26	29119627	Genotyping	German	CRC
27	28915636	Genotyping	Chinese	CRC
28	28039327	Genotyping	Chinese	CRC
29	27746584	Genotyping	Saudi Arabia	CRC
30	27665685	Genotyping	Chinese	CRC
31	27354594	Genotyping	Korea	CRC
32	26547791	Genotyping	Chinese	CRC.
33	26148620	GWAS	Chinese	CRC
34	26083022	Genotyping	Tunisian	CRC
35	25834816	Genotyping	Chinese	CRC
36	25680555	Genotyping	Iran	CRC
37	25640388	GWAS	Iran	CRC
38	25222241	Genotyping	Chinese	CRC
39	24777809	Genotyping	Chinese	CRC
40	24562971	Genotyping	African Americans	CRC

Abbreviations: ChIP-Seq: Chromatin immunoprecipitation-sequencing, CRC: Colorectal cancer, DNase-seq: DNase I hypersensitive sites sequencing, GWAS: Genome-wide association study,

WES: Whole exome sequencing, NHGRI:National Human Genome Research Institute, PMID: PubMed IDentifier. PMID obtained from PubMed database https://pubmed.ncbi.nlm.nih.gov/.

Table 2. Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000167207	NOD2	1	1	1	1	1	1	6
ENSG00000076242	MLH1	1	1	1	1	1	0	5
ENSG00000105329	TGFB1	1	1	1	1	1	0	5
ENSG00000141510	TP53	1	1	1	1	1	0	5
ENSG00000112116	IL17F	1	0	0	1	1	1	4
ENSG00000113318	MSH3	1	1	0	1	1	0	4
ENSG00000135862	LAMC1	1	1	0	1	1	0	4
ENSG00000073756	PTGS2	0	1	1	1	1	0	4
ENSG00000088305	DNMT3B	0	1	0	1	1	1	4
ENSG00000095303	PTGS1	0	1	1	1	1	0	4
ENSG00000112115	IL17A	0	1	1	1	1	0	4
ENSG00000132781	MUTYH	1	0	1	1	1	0	4
ENSG00000134899	ERCC5	1	1	0	1	1	0	4
ENSG00000143799	PARP1	0	1	1	1	1	0	4
ENSG00000149485	FADS1	0	1	1	1	1	0	4
ENSG00000150093	ITGB1	0	1	1	1	1	0	4
ENSG00000164362	TERT	0	0	1	1	1	1	4
ENSG00000183765	CHEK2	1	0	1	1	1	0	4

Table 2 (Cont). Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000232810	TNF	0	1	1	1	1	0	4
ENSG00000111424	VDR	1	1	1	1	0	0	4
ENSG00000039068	CDH1	0	1	0	1	1	0	3
ENSG00000082701	GSK3B	0	1	0	1	1	0	3
ENSG00000099194	SCD	0	1	0	1	1	0	3
ENSG00000101665	SMAD7	1	0	0	1	1	0	3
ENSG00000112715	VEGFA	0	1	0	1	1	0	3
ENSG00000117394	SLC2A1	0	1	0	1	1	0	3
ENSG00000118971	CCND2	0	1	0	1	1	0	3
ENSG00000123268	ATF1	0	1	0	1	1	0	3
ENSG00000130702	LAMA5	0	1	0	1	1	0	3
ENSG00000134824	FADS2	0	1	0	1	1	0	3
ENSG00000146674	IGFBP3	1	0	0	1	1	0	3
ENSG00000148737	TCF7L2	1	0	0	1	1	0	3
ENSG00000152270	PDE3B	0	1	0	1	1	0	3
ENSG00000168496	FEN1	0	1	0	1	1	0	3
ENSG00000259207	ITGB3	0	1	0	1	1	0	3
ENSG00000007312	CD79B	0	1	0	1	0	1	3

Table 2 (Cont). Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000100714	MTHFD1	1	1	0	1	0	0	3
ENSG00000106366	SERPINE1	0	0	1	1	1	0	3
ENSG00000114487	MORC1	1	1	0	1	0	0	3
ENSG00000121879	PIK3CA	0	0	1	1	1	0	3
ENSG00000124275	MTRR	1	1	0	1	0	0	3
ENSG00000134982	APC	1	0	0	1	1	0	3
ENSG00000136936	XPA	0	0	1	1	1	0	3
ENSG00000154767	XPC	0	1	0	1	1	0	3
ENSG00000172115	CYCS	0	1	0	1	1	0	3
ENSG00000189403	HMGB1	0	1	0	1	1	0	3
ENSG00000213886	UBD	1	1	0	1	0	0	3
ENSG00000108576	SLC6A4	0	1	1	1	0	0	3
ENSG00000140465	CYP1A1	1	0	1	1	0	0	3
ENSG00000146477	SLC22A3	0	1	1	1	0	0	3
ENSG00000148795	CYP17A1	0	1	0	1	0	0	2
ENSG00000162552	WNT4	0	0	0	1	1	0	2
ENSG00000162594	IL23R	0	0	0	1	1	0	2
ENSG00000196611	MMP1	0	0	0	1	1	0	2

Table 2 (Cont). Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000204414	CSHL1	1	1	0	0	0	0	2
ENSG00000047932	GOPC	0	1	0	1	0	0	2
ENSG00000050405	LIMA1	0	1	0	1	0	0	2
ENSG00000062038	CDH3	0	1	0	1	0	0	2
ENSG00000077514	POLD3	0	0	0	1	1	0	2
ENSG00000104918	RETN	0	1	0	1	0	0	2
ENSG00000111669	TPI1	0	1	0	1	0	0	2
ENSG00000112561	TFEB	0	1	0	1	0	0	2
ENSG00000122870	BICC1	1	0	0	1	0	0	2
ENSG00000126653	NSRP1	0	1	0	1	0	0	2
ENSG00000134897	BIVM	0	1	0	1	0	0	2
ENSG00000135111	TBX3	0	0	0	1	1	0	2
ENSG00000136937	NCBP1	0	1	0	1	0	0	2
ENSG00000142046	TMEM91	1	1	0	0	0	0	2
ENSG00000164307	ERAP1	0	1	0	1	0	0	2
ENSG00000170860	LSM3	0	1	0	1	0	0	2
ENSG00000171858	RPS21	0	1	0	1	0	0	2
ENSG00000173559	NABP1	0	1	0	1	0	0	2

Table 2 (Cont). Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000178567	EPM2AIP1	0	1	0	1	0	0	2
ENSG00000186104	CYP2R1	0	1	0	1	0	0	2
ENSG00000188641	DPYD	1	0	0	1	0	0	2
ENSG00000196793	ZNF239	1	1	0	0	0	0	2
ENSG00000199071	MIR423	0	1	0	0	1	0	2
ENSG00000212993	POU5F1B	0	1	0	0	1	0	2
ENSG00000228242	AC093495.4	1	1	0	0	0	0	2
ENSG00000060237	WNK1	0	0	1	1	0	0	2
ENSG00000144852	NR1I2	0	1	1	0	0	0	2
ENSG00000169567	HINT1	0	0	1	1	0	0	2

Abbreviations: Missense: Missense mutation; Cis-eQTL: Cis expression quantitative trait loci, KO mice: Knockout mouse phenotype; PPI: Protein-protein interaction; KEGG: Kyoto Encyclopedia of Genes and Genomes; PID: Primary immunodeficiency.

Table 3. Prioritization of drug repositioning candidates for the treatment of colorectal cancer (CRC) based on CMap comparison to capecitabine.

Drug	Original Use	Mode of Action	Target Drug	CMap (Score)	ID
Celecoxib	Pain	Antagonist	PTGS1/PTGS2	99.09	7541
Rucaparib	Advanced Ovarian Cancer	Antagonist	PARP	94.14	0311
Ibuprofen	Pain	Antagonist	PTGS1/PTGS2	92.25	5518
Naproxen	Pain	Antagonist	PTGS1/PTGS2	86.40	9232
Balsalazide	Pain	Antagonist	PTGS1/PTGS2	81.55	0256
Aspirin	Pain	Antagonist	PTGS1/PTGS2	80.06	3652
Sulindac	Pain	Antagonist	PTGS1/PTGS2	77.92	6108
Sertraline	Antidepresant	Antagonist	SLC6A4	74.63	6761
Abiraterone	Prostate Cancer	Antagonist	CYP17A1	72.01	1415
Calcitriol	Hypocalcemia	Agonist	VDR	65.92	7315
Fluoxetine	Antidepresant	Antagonist	SLC6A4	61.71	9102
Olaparib	Metastatic Breast Cancer (MBC)	Antagonist	PAPR	60.14	3016
Rofecoxib	Pain	Antagonist	PTGS2	61.11	3600
Diclofenac	Pain	Antagonist	PTGS1/PTGS2	61.38	2256
Amitriptyline	Antidepresant	Antagonist	SLC6A4	56.96	7926
Parecoxib	Pain	Antagonist	PTGS2	50.75	0121
Paroxetine	Antidepresant	Antagonist	SLC6A4	45.72	1163
Dasatinib	Chronic Myelogenous leukemia (CML)	Antagonist	CCND2	45.33	8571

Abbreviations: Connectivity Map (CMap), the results were ranked by CMap score (ranging -100 to 100); higher scores denote higher similarities with capecitabine (a standard drug for CRC). CMap scores were obtained from the 'CMap Touchstone' tool in the CLUE website (https://clue.io/).

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Pharmacological Research 161 (2020) 105203

Contents lists available at ScienceDirect

Pharmacological Research

journal homepage: www.elsevier.com/locate/yphrs

Integration of genetic variants and gene network for drug repurposing in colorectal cancer

Lalu Muhammad Irham [a,c], Henry Sung-Ching Wong [a,b], Wan-Hsuan Chou [a,b], Wirawan Adikusuma [a,d], Eko Mugiyanto [a,e], Wan-Chen Huang [f,h,*], Wei-Chiao Chang [a,b,g,h,i,*]

[a] Department of Clinical Pharmacy, School of Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, 11031, Taiwan
[b] Master Program for Clinical Pharmacogenomics and Pharmacoproteomics, School of Pharmacy, Taipei Medical University, Taipei, 11031, Taiwan
[c] Faculty of Pharmacy, University of Ahmad Dahlan, Yogyakarta, 55164, Indonesia
[d] Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, 83127, Indonesia
[e] Department of Pharmacy, Faculty of Health Science, University of Muhammadiyah Pekajangan Pekalongan, 51172, Indonesia
[f] Institute of Cellular and Organismic Biology, Academia Sinica, Taipei City 115, Taiwan
[g] Integrative Research Center for Critical Care, Wan Fang Hospital, Taipei Medical University, Taipei, 11696, Taiwan
[h] Department of Medical Research; Department of Pharmacy; Shuang Ho Hospital, Taipei Medical University, New Taipei City, 23561, Taiwan
[i] TMU Research Center of Cancer Translational Medicine, Taipei, Taiwan

ARTICLE INFO

Keywords:
Drug discovery
Drug repurposing
Data mining
Colorectal cancer
Bioinformatics

ABSTRACT

Even though many genetic risk loci for human diseases have been identified and comprehensively cataloged, strategies to guide clinical research by integrating the extensive results of genetic studies and biological resources are still limited. Moreover, integrative analyses that provide novel insights into disease biology are expected to be especially useful for drug discovery. Herein, we used text mining of genetic studies on colorectal cancer (CRC) and assigned biological annotations to identified risk genes in order to discover novel drug targets and potential drugs for repurposing. Risk genes for CRC were obtained from PubMed text mining, and for each gene, six functional and bioinformatic annotations were analyzed. The annotations include missense mutations, cis-expression quantitative trait loci (cis-eQTL), molecular pathway analyses, protein-protein interactions (PPIs), a genetic overlap with knockout mouse phenotypes, and primary immunodeficiency (PID). We then prioritized the biological risk candidate genes according to a scoring system of the six functional annotations. Each functional annotation was assigned one point, and those genes with a score $\geq$2 were designated "biological CRC risk genes". Using this method, we revealed 82 biological CRC risk genes, which were mapped to 128 genes in an expanded PPI network. Further utilizing DrugBank and the Therapeutic Target Database, we found 21 genes in our list that are targeted by 166 candidate drugs. Based on data from ClinicalTrials.gov and literature review, we found four known target genes with six drugs for clinical treatment in CRC, and three target genes with nine drugs supported by previous preclinical results in CRC. Additionally, 12 genes are targeted by 32 drugs approved for other indications, which can possibly be repurposed for CRC treatment. Finally, analysis from Connectivity Map (CMap) showed that 18 drugs have a high potential for CRC.

Abbreviations: BRAF, B-Raf proto-oncogene; CCND2, Cyclin-D2; CRC, Colorectal cancer; CMap, Connectivity map; COX-1, Cyclooxygenase-1; COX-2, Cyclo-oxygenase-2; CML, Chronic myeloid leukemia; Cis-eQTL, Cis-expressionquantitative trait loci; CYP1A1, Cytochrome P450 Family 1 Subfamily A Member 1; FDA, Food and Drug Administration; FAP, Familial adenomatous polyposis; GWAS, Genome-wide association study; GO, Gene ontology; HER-2, Human epidermal growth factor receptor 2; IUIS, International Union of Immunological Societies; IL17F, Interleukin 17F; KEGG, Kyoto Encyclopedia of Genes and Genomes; KRAS, Kirsten rat sarcoma; LD, Linkage disequilibrium; MLH1, MutL homolog 1; MSI, Microsatellite instability; NHGRI, National Human Genome Research Institute; NOD2, Domain-containing protein 2; NSAIDs, Nonsteroidal anti-inflammatory drugs; NRAS, Neuroblastoma RAS; PARP, Poly (ADP-ribose) polymerase; PMID, PubMed IDentifier; PPI, Protein-protein interaction; PID, Primary immunodeficiency; PI3K, Phosphatidylinositol 3-kinase; PIK3CA, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha; SNPs, Single nucleotide polymorphisms; SSRIs, Selective serotonin-reuptake inhibitors; TGFB1, Transforming growth factor beta-1; TP53, Tumor protein p53; VEGF, Vascular endothelial growth factor; VDR, Vitamin D receptor; WES, Whole exome sequencing.

* Corresponding authors.
E-mail addresses: wchuang1@gate.sinica.edu.tw (W.-C. Huang), wcc@tmu.edu.tw (W.-C. Chang).

https://doi.org/10.1016/j.phrs.2020.105203
Received 6 June 2020; Received in revised form 27 August 2020; Accepted 9 September 2020
Available online 17 September 2020

1. Introduction

Colorectal cancer (CRC) is the second most common cause of cancer morbidity and mortality, and it occurs as the fourth most frequent in males and the third most frequent in females globally [1]. Transformation of CRC is the result of a progressive accumulation of genetic and epigenetic alterations. Once the CRC is diagnosed, patient outcomes are largely dependent on the disease stage, with early-stage disease generally showing better outcomes after treatment, and advanced disease characterized as clinically aggressive [2]. Thus, more effective anticancer drugs are required, as treatment approaches have been limited [2]. Bringing a new drug through clinical trials to the market is time consuming and expensive, often requiring ~15 years and > US$1 billion [3]. Currently, more than 10,000 clinical trials for cancer are registered at www.clinicaltrials.gov, however, only a few drug candidates continue to subsequent phases [4]. In the end, only around 5 % of new molecules entering phase I clinical trials will eventually be approved by the U.S. Food and Drug Administration (FDA) for disease treatment [5]. Thus, the approach of drug repurposing, which consists of finding new uses for old drugs as therapeutic agents, has become increasingly popular. Advantages of drug repurposing include clear understandings of the drug's safety, pharmacological mechanisms and pharmacokinetic profiles [6]. One example of clinically successful drug repurposing is the use of aspirin for cardiovascular disease and CRC prevention; aspirin was initially used for pain relief [7,8]. Additionally, sildenafil was originally used to treat high blood pressure, but was repurposed for erectile dysfunction [9]. Therefore, drug repurposing based on genomic information may be a promising strategy to improve the efficiency of drug discovery process.

In order to find candidates for drug repurposing, the relationships between diseases and genes, genes and drugs, and diseases and drugs should be identified. One strategy to identify these relationships is literature searching for pathway prioritization [10]. For example, candidate genes can be prioritized using keywords to perform text mining on a database.The database contains a large-scale genomic information that may shed light on germline risk alleles in patients with disease [10,11]. In addition, the PubMed database contains more than 24 million citations from MEDLINE and other biomedical literature data sources [12]. In our study, we queried the PubMed database to identify candidate risk genes. Then, we utilized six bioinformatics repositories to identify the genes that most likely to have biological impact in CRC etiology, and we also identified relationships between those CRC genes and existing drugs. The current study design was adopted from Okada et al., who predicted drugs for repurposing to rheumatoid arthritis by utilizing risk genes of rheumatoid arthritis [13]. Notably, the genomic information from genome-wide association study (GWAS) was utilized by an *in silico* pipeline to study drug repurposing for type 2 diabetes and cancer [14,15].

Germline mutations that alter risks of CRC have been previously identified and used to search for target drugs [16,17]. However, successful translation of this basic research into clinical practice has been limited, and remains as a future challenge. Therefore, in this study, we determined candidate genes associated with CRC and further prioritized the candidate genes using a scoring system based on six *in silico* criteria. As a result, we identified candidate biological risk genes for CRC that provide information to guide the selection of drugs for repurposing to CRC.

2. Materials and methods

Integrative analyses of genetics-driven genomic drug repurposing for CRC by utilizing multiple databases are shown in Fig. 1. Single nucleotide polymorphism (SNPs) associated with CRC (CRC-associated SNPs) were obtained from PubMed text mining and expanded using HaploReg (v4.1) with a criterion of $r^2 > 0.8$. The genes corresponding to CRC-associated SNPs are denoted as "CRC-associated genes". Subsequently, genomic data were prioritized based on six functional annotation criteria. Those genes with one functional annotation were awarded one point (score) and those genes with a score ≥2 were classified as "biological CRC risk genes". The biological CRC risk genes were utilized in further analyses, including the use of the STRING database [18] to expand the list of candidate genes as the drug target genes. We mapped the expanded list of genes according to targets of approved drugs which found in the DrugBank and Therapeutic Target Database. The drug targets were examined to determine their clinical status according to ClinicalTrials.gov. Finally, we used the CMap database to prioritize the most promising candidate drugs for repurposing to CRC treatment.

2.1. Literature review by PubMed text mining

CRC-associated SNPs were identified through a systematic search of the PubMed database (articles published 2014–2018). The search terms used were: *"colorectal cancer", "risk gene", "susceptibility", "polymorphism", "SNP", and "genotype"*. Inclusion of data was based on the following criteria: 1) all samples were human samples; 2) the clinical phenotype was "CRC"; and 3) the identified SNPs were associated with CRC.

2.2. CRC–associated genes

CRC-associated SNPs are extended to the SNPs in high LD ($r^2 > 0.8$ in Asian population based on 1000 genome project) with the SNPs

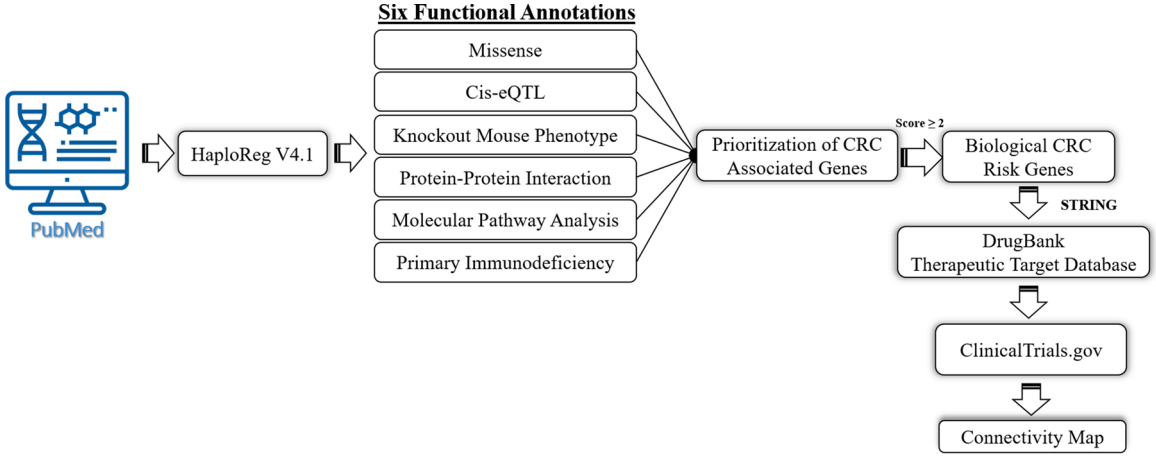


Fig. 1. This schematic model illustrates that genomic-based information cane be integrated for colorectal cancer (CRC) drug repurposing.

identified in the previous step using HaploReg v4.1 [19]. The LD criterion of $r^2 > 0.8$ was adopted from the study of Okada et al [13].

2.3. Functional annotation of CRC-associated genes

In order to prioritize the genes according to their biological function, six annotations were used to construct a scoring system that reflects the most likely candidate genes as CRC targets. The first annotation was missense or nonsense mutation according to HaploReg v4.1, which contains annotations of the functional consequences from a database of (db)SNPs [20]. HaploReg v4.1 also links genetic variants to cis-expression quantitative trait loci (cis-eQTL). Moreover, we examined the cis-eQTL in target tissues of whole blood or colon [20]. After that, in order to understand the connections between mutant genes and phenotypes, we employed WebGestalt 2017 to perform functional enrichment analyses [21]. Mammalian Phenotype (MP) Ontology, which contains information on phenotypes of mice and other mammals, was used [22]. To query the mouse phenotype, we converted genes from the human Ensemble ID to mouse Ensemble ID using BioMart [23]. Genes enriched in mouse phenotypes with a false discovery rate (*FDR*) of $<$ 0.05 were considered significant results. Furthermore, protein-protein interactions (PPIs) were used to identify CRC-associated genes enriched in biological protein networks [24]. WebGestalt 2017 was also used to perform enrichment analyses and investigate if the genes were enriched in any specific functional pathway [25]. Specifically, the biological process Gene Ontology (GO) categories in WebGestalt 2017 were interrogated for this step. Significance of a result was set at $FDR < 0.05$ [25]. To determine what kind of molecular pathways were enriched on the CRC-associated gene list and which gene was involved, enrichment analyses was performed on molecular pathways utilizing the Kyoto Encyclopedia of Genes and Genomes (KEGG), an online pathway database of biochemicals from WebGestalt 2017 [25]. Primary immunodeficiency (PID) genes are the last annotation criterion. PID diseases are innate immune diseases [26] and are often reported to associate with cancer [27]. PID diseases are known as a genetic disorders associated with an increased incidence of malignancies, such as lymphoproliferative disorders, gastric adenocarcinoma, and CRC [28].]. A hypergeometric test was used to perform an enrichment analysis on these data; the criterion for significance was p-value < 0.05.

2.4. Prioritization of biological candidate genes

According to these functional annotations, a score was assigned to each gene. One point was awarded for each criterion, as follows: 1) genes with missense or nonsense variants; (2) cis-eQTL genes of risk SNPs; (3) genes prioritized by the Knockout Mouse Phenotype; (4) genes prioritized by PPI network; (5) Molecular Pathway; (6) PID. Each gene was scored according to the number of criteria that were satisfied (scores ranged 0–6 for each gene). Genes with a score of $\geq$2 were categorized as 'biological CRC risk genes'. We set the threshold of biological score ≥ 2 to find much higher number of the biological CRC risk genes.

2.5. STRING database

The purpose of the STRING database (http://string-db.org/) is to identify functional interactions related to protein expression by integrating predicted protein-protein association data [18]. A large number of disease-protein networks may provide targets for diseases [29]. The proteins in direct PPI with these encoded by biological CRC risk genes were identified using STRING database version 10.5 (https://string-db.org/cgi/network,). The genes encoding these protein were defined as "drug target gene". Protein products from the identified biological CRC risk genes, or several genes from a direct PPI network, were explored as potential targets of approved drugs, either for CRC or other indications.

2.6. Drug validation and discovery

The DrugBank database (https://www.drugbank.ca/) and Therapeutic Targets Database (http://bidd.nus.edu.sg/BIDD-Databases/TTD/TTD.asp) were utilized to evaluate potential drug targets. Drug target genes were used to interrogate the databases according to several criteria, such as drugs with pharmacological activities, effectiveness in humans, and annotations of 'approved', 'clinical trial' or 'experimental drugs'. The identified drugs were reviewed on ClinicalTrials.gov (https://clinicaltrials.gov) to identify any clinical investigations for CRC or other diseases.

2.7. Connectivity map (CMap) analyses

To prioritize the list of drugs for CRC repurposing, we used the CMap touchstone database platform to rank drugs according to a connectivity score (-100 to 100) [30]. The CMap scores were obtained from the 'Touchstone' tool in the CLUE website (https://clue.io/). The CMap database can be used to perform profile similarity analyses, which is based on a collection of genome-wide transcriptional expression profiles from cultured human cancer cells treated with different drug compounds [31]. In this study, drugs were compared with capecitabine as standard treatment for CRC according to the National Comprehensive Cancer Network (NCCN) guideline [32].

2.8. Statistical analyses

All analytic workflows were performed using R Studio v3.4.3 and the haploR package (https://www.r-project.org/ and https://cran.r-project.org/web/packages/haploR/index.html). Over-representation analysis (ORA) was used to prioritize the genes in Knockout Mouse Phenotype, PPI network and Molecular Pathway. The *FDR* of $<$0.05 was determined to indicate the statistical significance. A hypergeometric test was used to perform an enrichment analysis for PID; the criterion for significance was p-value < 0.05.

3. Results

We retrieved SNPs associated with CRC from PubMed text mining. Forty CRC-associated studies were collected from genotyping studies and GWAS (Table 1). In total of 170 unique SNPs were found to associate with CRC risk (**Table S1**). The number of SNPs was expanded based on LD, using HaploReg v4.1 with criterion of $r^2 > 0.8$, and this expansion yielded 210 CRC-associated genes **(Table S2)**.

3.1. Functional annotations of CRC risk SNPs

In order to prioritize genes according to known biological processes, we devised six biological functional annotations based on 210 CRC-associated genes and assigned one point for each functional annotation. The results of six biological functional annotations as shown in **Table S2** and Fig. 2; we determined that 29 out of 210 genes (13.80 %) were missense or nonsense variants. We then assessed the cis-eQTL effects with HaploReg v4.1, using this strategy, we found 101 genes (48.09 %) risk SNPs have cis-eQTL effects on 210 genes, in either whole blood or colon tissues. Phenotype data were then retrieved from MP Ontology, which contains information on phenotypes of mice and other mammals. WebGestalt 2017 was used to perform an ORA, and 24 genes (11.43 %) were found to overlap with CRC risk genes ($FDR < 0.05$). Gene Ontology (GO) annotations obtained from WebGestalt 2017 were then used to evaluate PPIs. According to the PPIs, 95 genes (45.24 %) that overlapped with other CRC risk genes ($FDR < 0.05$) were found. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to perform an ORA on molecular pathways. A total of 49 CRC-associated genes (23.33 %) in KEGG pathways were successfully identified by the enrichment analysis. We also analyzed PID data from the IUIS and subsequently confirmed

Table 1
Profiles of colorectal cancer (CRC) reports in this study.

No	Type of the Study	Population	Source of Study (PMID)
1	WES	Chinese	30018674
2	GWAS	Italian	28915899
3	GWAS	European	24737748
4	Genotype	Saudi Arabian	25811207
5	GWAS	East Asian, Chinese, Japanese Korean	26965516
6	GWAS	East Asian	24836286
7.	Genotyping	Saudi Arabian	30233234
8.	GWAS	Chinese	26515597
9	GWAS	Japanese	25105248
10	Genotyping	Chinese	27698911
11	GWAS	Chinese	27145994
12	GWAS	Chinese	28218435
13	ChIP-Seq, DNase-Seq	Chinese	29267898
14	GWAS	East Asian	24448986
15	Genotyping	Korean	23875689
16	GWAS	European-Asian	23946381
17	GWAS	European	25990418
18	Genotyping	Chinese	29428571
19	Genotyping	Taiwanese	24968322
20	Genotyping	Tunisians	29689450
21	Genotyping	Iran	29953646
22	Genotyping	European	29372689
23	Genotyping	Chinese	29419695
24	GWAS	NHGRI	27146020
25	Genotyping	Chinese	29766219
26	Genotyping	German	29119627
27	Genotyping	Chinese	28915636
28	Genotyping	Chinese	28039327
29	Genotyping	Saudi Arabia	27746584
30	Genotyping	Chinese	27665685
31	Genotyping	Korea	27354594
32	Genotyping	Chinese	26547791
33	GWAS	Chinese	26148620
34	Genotyping	Tunisian	26083022
35	Genotyping	Chinese	25834816
36	Genotyping	Iran	25680555
37	GWAS	Iran	25640388
38	Genotyping	Chinese	25222241
39	Genotyping	Chinese	24777809
40	Genotyping	African Americans	24562971

Abbreviations: ChIP-Seq: Chromatin immunoprecipitation-sequencing, CRC: Colorectal cancer, DNase-seq: DNase I hypersensitive sites sequencing, GWAS: Genome-wide association study, WES: Whole exome sequencing, NHGRI: National Human Genome Research Institute.

overlapping genes with our findings. Five genes (2.38 %) with statistically significant overlap in PID ($p < 0.05$) were identified. After compiling the scores from 0-6, we found that genes with a score of 0 and 1 have the same numbers of genes, 64 genes, respectively. Genes with a score of 2 were 32 while those with a score of 3 were 30. There were 16 genes with a score of 4, three genes with a score of 5, and one gene with a score of 6 (Fig. 3). In total, we obtained eighty-two genes with a score of ≥2, which were categorized as 'biological CRC risk genes' **(**Table 2**)**. Here are the top five of biological CRC risk genes which were domain-containing protein 2 (NOD2), followed by MutL homolog 1 (MLH1), transforming growth factor beta-1 (TGFB1), tumor protein p53 (TP53) and interleukin 17 F (IL17 F).

3.2. Expanding the list of biological CRC risk genes

We used the STRING database of PPIs to expand the number of biological CRC risk genes. After expansion with the STRING database, 128 genes were included in the list (**Table S3**). These genes comprised the final list of candidate genes, which were used for futher analysis.

3.3. Discovery of CRC drug targets

We obtained 1108 gene pairs from curated PPI networking (**Table S4**) and found21 target genes in DrugBank and the Therapeutic Target Database (**Table S5**). The 166 drugs are either approved, in clinical trials, or experimental drugs for human diseases (**Table S6**). Four target genes with six drugs available used for CRCwere identified: genistein, marimastat and bevacizumab, dalteparin, aflibercept and tegafur uracil (Fig. 4). Importantly, three drugs (bevacizumab, aflibercept and tegafur uracil) are the most frequently used as chemotheraputic drugs for the treatment of CRC. Further, correlations of drugs approved for other diseases with biological CRC risk genes were also assessed. We found 12 known targets with 32 drugs approved for other diseases, but areunder clincal investigation for CRC now.These drugs may have potential to be repositioned for treatment of CRC (Fig. 5**)**. An example is dasatinib, which is approved for treating chronic myeloid leukemia (CML), but has been repurposed for CRC treatment in clinical trials. Another drug, genistein, cytochrome P450 family 1 subfamily a member 1 (CYP1A1) targeting agent, is also a promising drug for treatment of CRC. Genistein may inhibit cancer cell growth by blocking Wnt signaling required for CRC cell growth **(Table S7**). In addition, we also identified five non-steroidal anti-inflammatory drugs (NSAIDs) and four selective serotonin reuptake inhibitors (SSRIs), which have been widely reported in the preclinical *in vitro* or *in vivo* (CRC model) investigations for repurposing to CRC (Fig. 6). Especially, NSAIDs include balsalazide, tolfenamic acid, adapalene, sulindac, and parecoxib

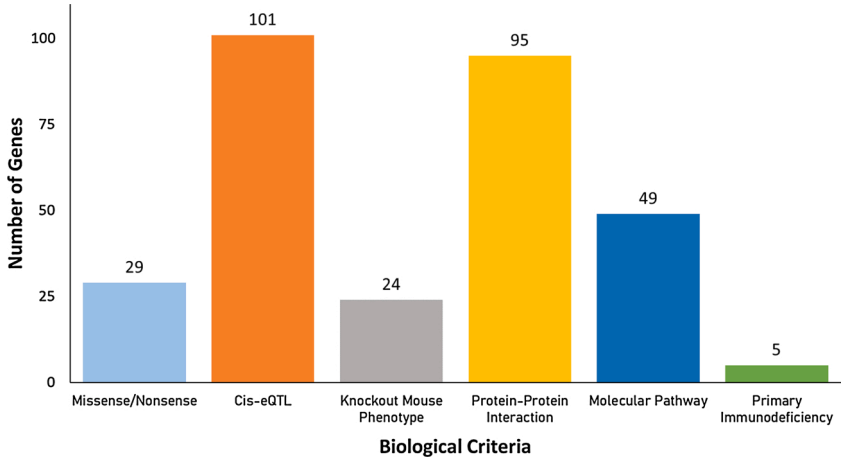


Fig. 2. Bar chart shows the number of genes (*y*-axis) satisfying each of the six biological criteria (*x*-axis) for drug prioritization.

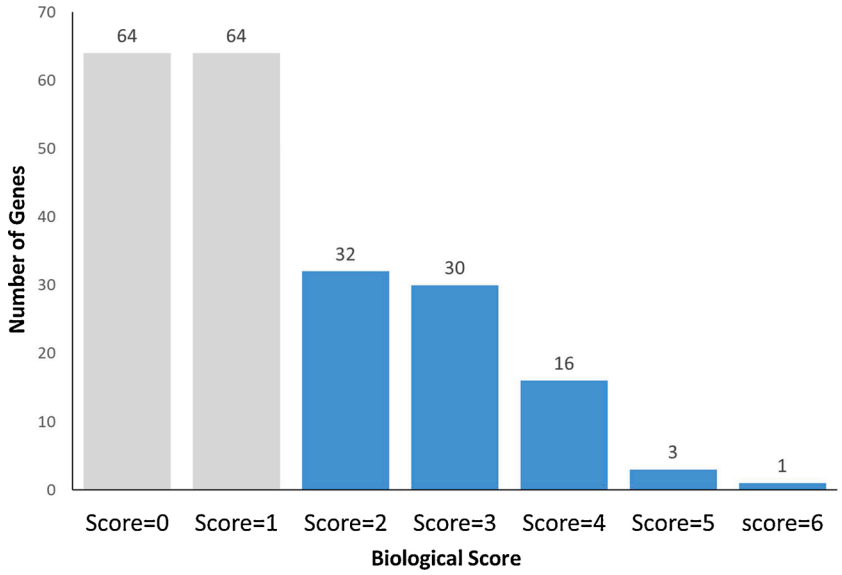


Fig. 3. Distribution plot shows the number of genes (*y*-axis) with different biological (gene) scores (*x*-axis). After compiling the scores, 82(=32+30+16+3+1) genes with a score of ≥2 were categorized as 'biological CRC risk genes' (blue).

Table 2
Prioritized biological annotations for colorectal cancer (CRC) genes with a score ≥ 2.

GENCODE_ID	GENCODE_name	Missense	Cis-eQTL	KO mice	PPI	KEGG	PID	Total Score
ENSG00000167207	NOD2	1	1	1	1	1	1	6
ENSG00000076242	MLH1	1	1	1	1	1	0	5
ENSG00000105329	TGFB1	1	1	1	1	1	0	5
ENSG00000141510	TP53	1	1	1	1	1	0	5
ENSG00000112116	IL17F	1	0	0	1	1	1	4
ENSG00000113318	MSH3	1	1	0	1	1	0	4
ENSG00000135862	LAMC1	1	1	0	1	1	0	4
ENSG00000073756	PTGS2	0	1	1	1	1	0	4
ENSG00000088305	DNMT3B	0	1	0	1	1	1	4
ENSG00000095303	PTGS1	0	1	1	1	1	0	4
ENSG00000112115	IL17A	0	1	1	1	1	0	4
ENSG00000132781	MUTYH	1	0	1	1	1	0	4
ENSG00000134899	ERCC5	1	1	0	1	1	0	4
ENSG00000143799	PARP1	0	1	1	1	1	0	4
ENSG00000149485	FADS1	0	1	1	1	1	0	4
ENSG00000150093	ITGB1	0	1	1	1	1	0	4
ENSG00000164362	TERT	0	0	1	1	1	1	4
ENSG00000183765	CHEK2	1	0	1	1	1	0	4

Abbreviations: Missense: Missense mutation; Cis-eQTL: Cis expression quantitative trait loci, KO mice: Knockout mouse phenotype; PPI: Protein-protein interaction; KEGG: Kyoto Encyclopedia of Genes and Genomes; PID: Primary immunodeficiency.

[33–36], and SSRI drugs include amitriptyline, fluoxetine, paroxetine and sertaline [37–40] (**Table S8**). Therefore, our results suggest that this combinational analysis of PubMed text-mining and six functional annotations can efficiently identify potential candidate drugs for CRC drug repurposing.

3.4. Connectivity Map (CMap) analyses prioritize the most promising drugs for CRC

To prioritize the most potential candidate drugs for CRC, we analyzed 41 drugs (32 drugs under clinical and 9 drugs preclinical investigation for CRC) using the CMap database. If two drugs have a strongly positive relationship, they may have similar effects in treating CRC. We found that 18 of 41 drugs had positive correlations with capecitabine, the standard treatment for CRC, according to the high CMap scores (Table 3). Among the 18 drugs, nine drugs (celecoxib, ibuprofen, naproxen, balsalazide, aspirin, sulindac, rofecoxib, diclofenac and parecoxib) target cyclooxygenase (COX)-1 and COX-2, three drugs (sertraline, fluoxetine and amitriptyline) target SLC6A4, two drugs (rucaparib and olaparib) target poly (ADP-ribose) polymerase (PARP), and one drug (dasatinib) targets cyclin-D2 (CCND2). As depicted in Table 3, the top five drugs that exhibited strong relationships with capecitabine were celecoxib, rucaparib, ibuprofen, naproxen and balsalazide, with respective CMap scores of 99.09, 94.14, 92.25, 86.40, and 81.55. This finding suggests the biological plausibility of repurposing cancer drugs from non-cancer drugs or other drugs for CRC

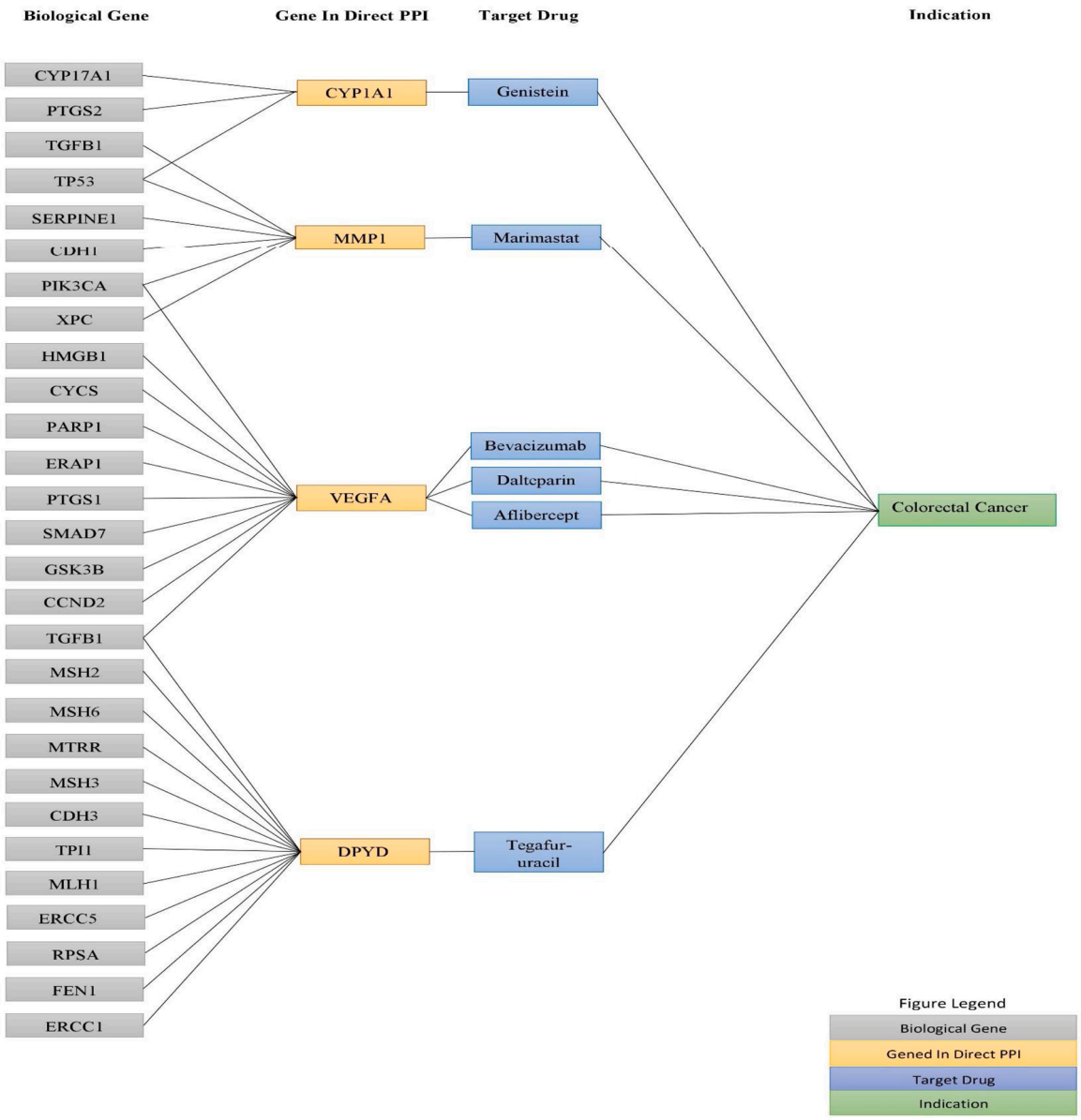


Fig. 4. Connections between biological CRC genes, genes from protein-protein ineteractions (PPIs, and drugs available for CRC.

treatment.

4. Discussion

In the current study, we mined PubMed texts to extract potential candidate genes that may guide drug repurposing to CRC. Six functional annotations were applied to build a scoring system and to prioritize CRC risk genes which are likely to be useful drug targets. We found 128 genes from the expanded PPI network, including four known targets of six clinical CRC drugs, 12 known targets for 32 drugs under clinical investigation for CRC. In addition, three known target genes assocaited with CRC can be modulated by nine drugs. Finally, by using the CMap database, we observed that18 drugs might have high potential to be repurposed for CRC. We also found six drugs, including genistein, marimastat, bevacizumab, dalteparin, aflibercept, and tegafur-uracil, that are currently in clinical trials for CRC. Genistein [NCT01985763] and Aflibercept [NCT03264274] are currently in a phase 2 clinical trial for treatment of metastatic CRC. Xeloda or UFT (tegafur-uracil) with folinic acid is in a phase 3 investigation for metastatic CRC [NCT00905047]. Bevacizumab is in a phase 2 clinical trial for advanced CRC in combination with irinotecan, lederfolin, and 5FU (FOLFIRI) [NCT01853813]. Dalteparin was assessed by a double-blinded phase 3 randomized control trial for treating advanced cancers, including breast, lung and colorectal [NCT00003674].

Although adjuvant therapy is available for patients with high-risk stage II and stage III CRC, around 40–50 % of patients experience recurrence with the development of metastatic disease [41]. For advanced CRC, targeting the vascular endothelial growth factor (VEGF) pathway has a pivotal role in increasing overall survival time in patients receiving chemotherapy [42]. Herein, VEGF is one of the drug targets in this study; VEGF can be targeted by bevacizumab, dalteparin, and aflibercept. Two of these drugs (bevacizumab and aflibercept) have been approved for use in metastatic CRC [43]. Indeed, most drugs that we found have indications for other cancer types. 32 drugs identified in this category are depicted in Fig. 5 (*e.g.*, ovarian cancer [rucaparib, niraparib, olaparib], non-small cell lung cancer [pemetrexed], prostate cancer [abiraterone], osteosarcoma [mifamurtide], CML [dasatinib], multiple myeloma [pomalidomide]). Another set of drugs identified in our study are currently used for autoimmune disease (*e.g.*, ankylosing spondylitis [etanercept, adalimumab, infliximab, golimumab,

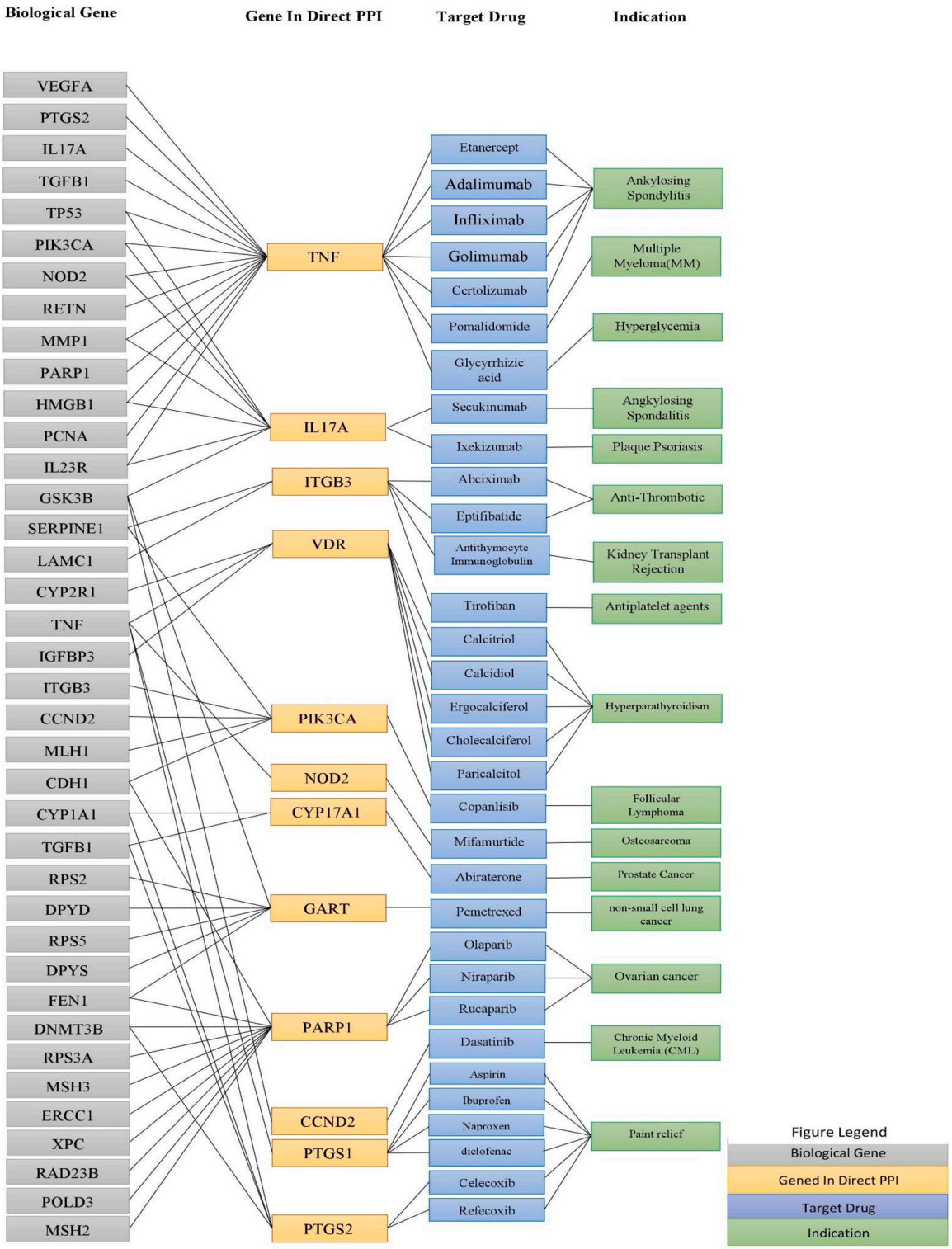


Fig. 5. Connections between biological CRC genes with the drugs that approved for other indications and are under clinical investigation for CRC.

secuximab]). Surprisingly, one drug we identified here as potentially useful for CRC is vitamin D. Vitamin D deficiency may affect various human cancers, and the Vitamin D receptor (VDR), which plays a role in regulating active vitamin D metabolites, is highly expressed in the small intestine and colon [44]. Vitamin D has been shown to protect from CRC tumorigenesis by binding long-chain fatty acids and bile acids in the small intestine and protecting colonic epithelial cells from mutagens [45]. Notably, Vitamin D-related drugs are currently being evaluated in CRC: cholecalciferol is in a phase 2 trial [NCT01074216]; cholecalciferol [NCT02172651] and paricalcitol [NCT01197664] are in phase 1 trials. Furthermore, we identified calcitriol as a promising drug for CRC repurposing based on the CMap analysis (Table 3). However, molecular evidence for the role of vitamin D in CRC is limited, so future functional studies are needed to investigate the mechanisms of how vitamin D affects CRC.

Another category of drugs identified in our study is NSAIDs, which

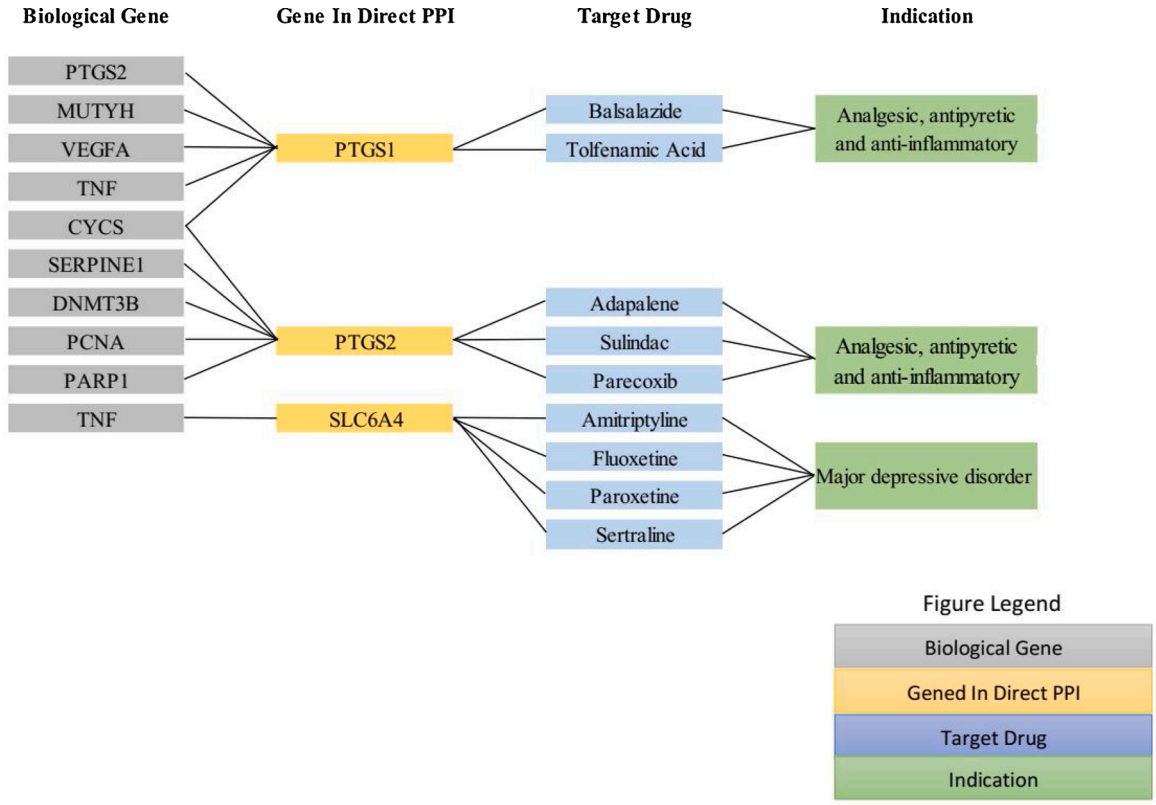


Fig. 6. Connections between biological CRC genes and drugs with preclinical data for CRC.

generally hold indications as anti-inflammatory agents [46]. From the results of a nationwide population-based case–control study, Kuo et al., indicated that NSAIDs use was associated with lower incidence of colorectal cancer [47]. According to the CMap analysis we performed in this study, nine NSAIDs (celecoxib, ibuprofen, naproxen, balsalazide, aspirin, sulindac, rofecoxib, diclofenac and parecoxib) were among the most promising for CRC repurposing. Furthermore, NSAIDs were also reported to reduce the growth of cancer cells through inhibition of two enzymes, COX-1 and COX-2 [48]. Aspirin is commonly used in cardiovascular disease treatment and to prevent myocardial infarction. Previous clinical studies indicated that post-diagnosis use of aspirin is correlated with favorable survival in CRC patients [49,50]. Ibuprofen was also reported to effectively inhibit the growth of prostate cancer [51]. In an *in vitro* study on adenocarcinoma gastric cells, the antitumor activity of ibuprofen was attributed to the induction of apoptosis and the reduction of cell proliferation, which attenuated angiogenesis [52]. Naproxen is a propionic-acid derivative and non-selective COX-inhibitor that was shown to inhibit cell proliferation, trigger apoptosis and suppress metastasis. *In vitro* and *in vivo* studies revealed that naproxen also had antineoplastic properties in leukemia, breast, colon, bladder, and osteosarcoma cell lines [53]. Furthermore, according to an *in vivo* study on a colon cancer animal model, the combination of naproxen and atorvastation significantly inhibited growth of colonic adenocarcinomas [54]. Diclofenac is a derivative of acetic acid that has a moderate selectivity for COX-2 inhibition. Diclofenac was reported to have antineoplastic effects in various tumor types, including fibrosarcomas, hepatomas, and colon, ovarian and pancreatic cancer [55]. Diclofenac also was reported to have anti-proliferative effects in human colon cancer cell lines [56]. In addition, other NSAIDs that we identified are celecoxib and rofecoxib for CRC. Celecoxib and rofecoxib are two COX-2 inhibitors approved by the U.S. FDA for FAP patients, and these drugs are also being studied in patients with sporadic adenomas and other cancer types [57]. A study of celecoxib as an adjuvant chemotherapy is ongoing in patients with metastatic CRC (phase 4) [NCT03645187]. Yang et al., reported that it may be effective for ovarian cancer [12]. Kim et al. suggested that the combination of celecoxib and paclitaxel could be an effective treatment for ovarian cancer [58]. Moreover, NSAIDs are being evaluated in preclinical models for repurposing to CRC (Fig. 6); such drugs including balsalazide, tolfenamic acid, adapalene and sulindac, which showned anticancer effects in human colon cancer cells [33–36].

Additionally, CMap analysis also showed that the SSRIs drugs, sertraline, fluoxetine and amitriptyline, may be promising candidates for repurposing to CRC. SSRI drugs have been widely used for the treatment of depression and anxiety [59]. Recently, several reports indicated that SSRI drugs have anticancer activity based on the inhibition of proliferation and induction of apoptosis in human colon cancer cells [37–40]. Stopper et al., reported that fluoxetine reduced tumor growth of colon cancer in animals and the incidence of colon cancer in humans through a blockage in tumor metabolism [38]. A following investigation also supported the anticancer activity of sertraline in the two human colon cancer cell lines (HT29 and LS1034) [40]. Currently, biomarkers for CRC are still limited to kirsten rat sarcoma (KRAS), neuroblastoma RAS (NRAS) and B-Raf proto-oncogene (BRAF), with the upcoming standard of care including phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA), human epidermal growth factor receptor 2 (HER-2) and microsatellite instability (MSI) [43]. PIK3CA is one of the therapeutic targets that we found in this study. PIK3CA is very likely to be targeted by copanlisib, which has an original indication for follicular lymphoma. Copanlisib, a phosphatidylinositol 3-kinase (PI3K) inhibitor, combined with nivolumab, an anti-PD-1 agent, is under evaluation in patients with mismatch-repair proficient (MSS) colorectal cancer [NCT03711058]. Furthermore, the safety and efficacy of copanlisib in combination with nivolumab are being assessed in an on-going phase 1b/2 study [NCT03735628].

The advantage of drug repurposing is that the safety data of identified drugs are already available [60]. Thus, *in silico* approach might be useful to narrow down candidate drugs for CRC disease, resulting in cost

Table 3
Prioritization of drug repositioning candidates for the treatment of colorectal cancer (CRC) based on CMap comparison to capecitabine.

Drugs	Original Use	Mode of Action	Target Drugs	CMap (Score)	ID
Celecoxib	Pain	Antagonist	PTGS1/ PTGS2	99.09	7541
Rucaparib	Advanced Ovarian Cancer	Antagonist	PARP	94.14	0311
Ibuprofen	Pain	Antagonist	PTGS1/ PTGS2	92.25	5518
Naproxen	Pain	Antagonist	PTGS1/ PTGS2	86.40	9232
Balsalazide	Pain	Antagonist	PTGS1/ PTGS2	81.55	0256
Aspirin	Pain	Antagonist	PTGS1/ PTGS2	80.06	3652
Sulindac	Pain	Antagonist	PTGS1/ PTGS2	77.92	6108
Sertraline	Antidepresant	Antagonist	SLC6A4	74.63	6761
Abiraterone	Prostate Cancer	Antagonist	CYP17A1	72.01	1415
Calcitriol	Hypocalcemia	Agonist	VDR	65.92	7315
Fluoxetine	Antidepresant	Antagonist	SLC6A4	61.71	9102
Olaparib	Metastatic Breast Cancer (MBC)	Antagonist	PAPR	60.14	3016
Rofecoxib	Pain	Antagonist	PTGS2	61.11	3600
Diclofenac	Pain	Antagonist	PTGS1/ PTGS2	61.38	2256
Amitriptyline	Antidepresant	Antagonist	SLC6A4	56.96	7926
Parecoxib	Pain	Antagonist	PTGS2	50.75	0121
Paroxetine	Antidepresant	Antagonist	SLC6A4	45.72	1163
Dasatinib	Chronic Myelogenous leukemia (CML)	Antagonist	CCND2	45.33	8571

Abbreviations: Connectivity Map (CMap), the results were ranked by CMap score (ranging -100 to 100); higher scores denote higher similarities with capecitabine (a standard drug for CRC). CMap scores were obtained from the 'CMap Touchstone' tool in the CLUE website (https://clue.io/).

and time savings. In particular, we would like to emphasize that CRC gene network -driven drug repurposing is a new model to contribute to CRC treatment.. However, our approaches have some limitations. The target genes that we identified are not all in pharmacological activities, therefore these genes might potentially miss the target of the drugs (undruggable). Our analysis showed that only 21 targets of the 1108 CRC target genes are druggable (genetic driven druggable). The comprehensive analysis on the druggability of genes performed by Finan et al. also revealed that only 4479 (22 %) among 20,300 protein coding genes are druggable [61]. Moreover, functional studies and clinical are still needed to confirm the safety and efficacy.

5. Conclusion

By utilizing genomic data to map disease-gene-protein-drug relationships, we found 82 biological CRC risk genes. Then, 128 genes were further expanded byPPI network. Among the identified targets, six drugs are clinically used for CRC, nine drugs were widely supported by preclinical data, and 32 drugs already approved for other indications might be good candidates for CRC treatment. Moreover, CMap results indicated that 18 drugs have potential to be repurposed for CRC. In conclusion, our study demonstrates the feasibility of utilizing genomic data and gene network as a potential method for drug discovery.

Author contributions

L.M.I., H.S.C.W., W.H.C and W.C.C conceived and designed the study. L.M.I performed the computational analysis. L.M.I wrote the manuscript. W.C.C provided the funding. H.S.C.W., W.H.C., W.C.H., W. A., E.M., and W.C.C revised the manuscript. W.C.H and W.C.C supervised and coordinated this study. All authors have read and approved of the manuscript and have made significant contributions to this study.

Data availability

The dataset used and/or analyzed in the study are obtained from ClinicalTrials.gov (https://clinicaltrials.gov/); CMap 'Touchstone' tool in the CLUE website (https://clue.io/); Drug bank databases (https://www.drugbank.ca/); HaploReg (v4.1) (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php); STRING database (http://string-db.org/); Therapeutic Targets Database (http://bidd.nus.edu.sg/BIDD-Databases/TTD/TTD.asp); WebGestalt 2017 (http://www.webgestalt.org/2017/option.php); PubMed database (https://pubmed.ncbi.nlm.nih.gov/).

Declaration of Competing Interest

The authors disclose no conflict of interest.

Acknowledgements

This project was supported by grants from the Ministry of Science and Technology, Taiwan (MOST105-2628-B-038-001-MY4, 109-2314-B-038-131) and TMU Research Center of Cancer Translational Medicine from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.phrs.2020.105203.

References

[1] R. Siegel, J. Ma, Z. Zou, et al., Cancer statistics, 2014, CA Cancer J. Clin. 64 (1) (2014) 9–29.
[2] H.J. Freeman, Early stage colon cancer, World J. Gastroenterol. 19 (46) (2013) 8468–8473.
[3] M. Sirota, J.T. Dudley, J. Kim, et al., Discovery and preclinical validation of drug indications using compendia of public gene expression data, Sci. Transl. Med. 3 (96) (2011), 96ra77.
[4] M. Hay, D.W. Thomas, J.L. Craighead, et al., Clinical development success rates for investigational drugs, Nat. Biotechnol. 32 (1) (2014) 40–51.
[5] S.M. Paul, D.S. Mytelka, C.T. Dunwiddie, et al., How to improve R&D productivity: the pharmaceutical industry's grand challenge, Nature reviews, Drug discovery 9 (3) (2010) 203–214.
[6] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, Nature reviews, Drug discovery 3 (8) (2004) 673–683.
[7] A.M. Algra, P.M. Rothwell, Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials, Lancet Oncol. 13 (5) (2012) 518–527.
[8] T. Wolff, T. Miller, S. Ko, Aspirin for the primary prevention of cardiovascular events: an update of the evidence for the U.S. Preventive Services Task Force, Ann. Intern. Med. 150 (6) (2009) 405–410.
[9] H.A. Ghofrani, I.H. Osterloh, F. Grimminger, Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond, Nature reviews, Drug discovery 5 (8) (2006) 689–702.
[10] Y. Luo, G. Riedlinger, P. Szolovits, Text mining in cancer gene and pathway prioritization, Cancer Inform. 13 (Suppl 1) (2014) 69–79.
[11] A. Singhal, M. Simmons, Z. Lu, Text mining genotype-phenotype relationships from biomedical literature for database curation and precision medicine, PLoS Comput. Biol. 12 (11) (2016), e1005017.
[12] H.T. Yang, J.H. Ju, Y.T. Wong, et al., Literature-based discovery of new candidates for drug repurposing, Brief. Bioinformatics 18 (3) (2017) 488–497.
[13] Y. Okada, D. Wu, G. Trynka, et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery, Nature 506 (7488) (2014) 376–381.
[14] M. Imamura, A. Takahashi, T. Yamauchi, et al., Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes, Nat. Commun. 7 (2016) 10531.
[15] J. Zhang, K. Jiang, L. Lv, et al., Use of genome-wide association studies for cancer research and drug repositioning, PLoS One 10 (3) (2015), e0116477.
[16] E.M. Stoffel, E. Koeppe, J. Everett, et al., Germline genetic features of young individuals with colorectal Cancer, Gastroenterology 154 (4) (2018) 897–905, e1.

[17] S.A. Smith, T. French, S.J. Hollingsworth, The impact of germline mutations on targeted therapy, J. Pathol. 232 (2) (2014) 230–243.
[18] D. Szklarczyk, J.H. Morris, H. Cook, et al., The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible, Nucleic Acids Res. 45 (D1) (2017) D362–d368.
[19] I.Y. Zhbannikov, K. Arbeev, S. Ukraintseva, et al., haploR: an R package for querying web-based annotation tools, F1000Research 6 (2017) 97.
[20] L.D. Ward, M. Kellis, HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease, Nucleic Acids Res. 44 (D1) (2016) D877–81.
[21] J. Wang, S. Vasaikar, Z. Shi, et al., WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit, Nucleic Acids Res. 45 (W1) (2017) W130–w137.
[22] C.L. Smith, J.T. Eppig, The mammalian phenotype ontology: enabling robust annotation and comparative analysis, Wiley interdisciplinary reviews, Systems biology and medicine 1 (3) (2009) 390–399.
[23] S. Durinck, Y. Moreau, A. Kasprzyk, et al., BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis, Bioinformatics (Oxford, England 21 (16) (2005) 3439–3440.
[24] J. De Las Rivas, C. Fontanillo, Protein-protein interactions essentials: key concepts to building and analyzing interactome networks, PLoS Comput. Biol. 6 (6) (2010), e1000807.
[25] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology, The Gene Ontology Consortium, Nature genetics 25 (1) (2000) 25–29.
[26] N. Parvaneh, J.L. Casanova, L.D. Notarangelo, et al., Primary immunodeficiencies: a rapidly evolving story, J. Allergy Clin. Immunol. 131 (2) (2013) 314–323.
[27] E. Mortaz, P. Tabarsi, D. Mansouri, et al., Cancers related to immunodeficiencies: update and perspectives, Front. Immunol. 7 (2016) 365.
[28] K. Salavoura, A. Kolialexi, G. Tsangaris, et al., Development of cancer in patients with primary immunodeficiencies, Anticancer Res. 28 (2b) (2008) 1263–1269.
[29] H. Ruffner, A. Bauer, T. Bouwmeester, Human protein-protein interaction networks and the value for drug discovery, Drug Discov. Today 12 (17–18) (2007) 709–716.
[30] A. Subramanian, R. Narayan, S.M. Corsello, et al., A next generation connectivity map: L1000 platform and the first 1,000,000 profiles, Cell 171 (6) (2017) 1437–1452, e17.
[31] J. Lamb, E.D. Crawford, D. Peck, et al., The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease, Science 313 (5795) (2006) 1929–1935.
[32] N.C.C. Network, NCCN Guideline for Colon Cancer, 2018, p. 26.
[33] H.-Y. Kim, S.-L. Kim, Y.-R. Park, et al., Balsalazide potentiates parthenolide-mediated inhibition of nuclear Factor-κB signaling in HCT116 human colorectal Cancer cells, Intest. Res. 13 (3) (2015) 233–241.
[34] S.-H. Lee, J.H. Bahn, C.K. Choi, et al., ESE-1/EGR-1 pathway plays a role in tolfenamic acid-induced apoptosis in colorectal cancer cells, Mol. Cancer Ther. 7 (12) (2008) 3739–3750.
[35] H.N. Tinsley, B.D. Gary, J. Thaiparambil, et al., Colon tumor cell growth-inhibitory activity of sulindac sulfide and other nonsteroidal anti-inflammatory drugs is associated with phosphodiesterase 5 inhibition, Cancer Prev. Res. Phila. (Phila) 3 (10) (2010) 1303–1313.
[36] W. Xiong, W.H. Li, Y.X. Jiang, et al., Parecoxib: an enhancer of radiation therapy for colorectal cancer, Asian Pac. J. Cancer Prev. 16 (2) (2015) 627–633.
[37] H. Arimochi, K. Morita, Characterization of cytotoxic actions of tricyclic antidepressants on human HT29 colon carcinoma cells, Eur. J. Pharmacol. 541 (1–2) (2006) 17–23.
[38] H. Stopper, S.B. Garcia, A.M. Waaga-Gasser, et al., Antidepressant fluoxetine and its potential against colon tumors, World J. Gastrointest. Oncol. 6 (1) (2014) 11–21.
[39] W.J. Jang, S.K. Jung, T.T.L. Vo, et al., Anticancer activity of paroxetine in human colon cancer cells: involvement of MET and ERBB3, J. Cell. Mol. Med. 23 (2) (2019) 1106–1115.
[40] I. Gil-Ad, A. Zolokov, L. Lomnitski, et al., Evaluation of the potential anti-cancer activity of the antidepressant sertraline in human colon cancer cell lines and in colorectal cancer-xenografted mice, Int. J. Oncol. 33 (2) (2008) 277–286.
[41] S. Gill, A.W. Blackstock, R.M. Goldberg, Colorectal cancer, Mayo Clin. Proc. 82 (1) (2007) 114–129.
[42] B. Gustavsson, G. Carlsson, D. Machover, et al., A review of the evolution of systemic chemotherapy in the management of colorectal cancer, Clin. Colorectal Cancer 14 (1) (2015) 1–10.
[43] P.J. Cuyle, H. Prenen, Current and future biomarkers in the treatment of colorectal cancer, Acta Clin. Belg. 72 (2) (2017) 103–115.
[44] J. Sun, The role of vitamin d and vitamin d receptors in Colon Cancer, Clin. Transl. Gastroenterol. 8 (6) (2017) e103-e103.
[45] S.A. Lamprecht, M. Lipkin, Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis, Ann. N. Y. Acad. Sci. 952 (2001) 73–87.
[46] J.K. Buer, Origins and impact of the term' NSAID', Inflammopharmacology 22 (5) (2014) 263–267.
[47] C.N. Kuo, J.J. Pan, Y.W. Huang, et al., Association between nonsteroidal anti-inflammatory drugs and colorectal Cancer: a population-based case-control study, Cancer epidemiology, Cancer Epidemiol. Biomarkers Prev. 27 (7) (2018) 737–745.
[48] J.Y. Wang, J. Sun, M.Y. Huang, et al., STIM1 overexpression promotes colorectal cancer progression, cell motility and COX-2 expression, Oncogene 34 (33) (2015) 4358–4367.
[49] M.S. Reimers, E. Bastiaannet, M.P. van Herk-Sukel, et al., Aspirin use after diagnosis improves survival in older adults with colon cancer: a retrospective cohort study, J. Am. Geriatr. Soc. 60 (12) (2012) 2232–2236.
[50] C. McCowan, A.J. Munro, P.T. Donnan, et al., Use of aspirin post-diagnosis in a cohort of patients with colorectal cancer and its association with all-cause and colorectal cancer specific mortality, Eur. J. Cancer 49 (5) (2013) 1049–1057.
[51] M.H. Kim, J. Chung, Synergistic cell death by EGCG and ibuprofen in DU-145 prostate cancer cell line, Anticancer Res. 27 (6b) (2007) 3947–3956.
[52] H. Akrami, S. Aminzadeh, H. Fallahi, Inhibitory effect of ibuprofen on tumor survival and angiogenesis in gastric cancer cell, Tumour Biol. 36 (5) (2015) 3237–3243.
[53] M.S. Kim, J.E. Kim, D.Y. Lim, et al., Naproxen induces cell-cycle arrest and apoptosis in human urinary bladder cancer cell lines and chemically induced cancers by targeting PI3K, Cancer Prev. Res. Phila. (Phila) 7 (2) (2014) 236–245.
[54] N. Suh, B.S. Reddy, A. DeCastro, et al., Combination of atorvastatin with sulindac or naproxen profoundly inhibits colonic adenocarcinomas by suppressing the p65/beta-catenin/cyclin D1 signaling pathway in rats, Cancer Prev. Res. Phila. (Phila) 4 (11) (2011) 1895–1902.
[55] L. Sleire, H.E. Forde, I.A. Netland, et al., Drug repurposing in cancer, Pharmacol. Res. 124 (2017) 74–91.
[56] A.V. Smirnova, L.B. Lazebnik, I.E. Trubitsina, et al., [Antiproliferative activity of diclofenac at tumor cell cultures], Eksp. Klin. Gastroenterol. 5 (2012) 66–69.
[57] A.J. Chu, T.H. Chou, B.D. Chen, Prevention of colorectal cancer using COX-2 inhibitors: basic science and clinical applications, Front. Biosci. 9 (2004) 2697–2713.
[58] H.J. Kim, G.W. Yim, E.J. Nam, et al., Synergistic effect of COX-2 inhibitor on paclitaxel-induced apoptosis in the human ovarian Cancer cell line OVCAR-3, Cancer Res. Treat. 46 (1) (2014) 81–92.
[59] J. Ramsberg, C. Asseburg, M. Henriksson, Effectiveness and cost-effectiveness of antidepressants in primary care: a multiple treatment comparison meta-analysis and cost-effectiveness model, PLoS One 7 (8) (2012) e42003-e42003.
[60] M.R. Hurle, L. Yang, Q. Xie, et al., Computational drug repositioning: from data to therapeutics, Clin. Pharmacol. Ther. 93 (4) (2013) 335–341.
[61] C. Finan, A. Gaulton, F.A. Kruger, et al., The druggable genome and support for target identification and validation in drug development, Sci. Transl. Med. 9 (383) (2017) eaag1166.