# Disease spectrum of gastric cancer susceptibility genes

**Cite this article as:** Sophia K. McKinley, Preeti Singh, Kanhua Yin, Jin Wang, Jingan Zhou, Yujia Bao, Menghua Wu, Kush Pathak, John T. Mullen, Danielle Braun and Kevin S. Hughes, Disease spectrum of gastric cancer susceptibility genes, Medical Oncology https://doi.org/10.1007/s12032-021-01495-w

# Disease Spectrum of Gastric Cancer Susceptibility Genes

Running title: Gastric cancer susceptibility genes

Sophia K. McKinley, MD, EdM[1]; Preeti Singh, MD[2]; Kanhua Yin, MD, MPH[2, 3]; Jin Wang, MD[2,4]; Jingan Zhou, MD[2,5]; Yujia Bao, MS[6]; Menghua Wu, BS[6]; Kush Pathak, MD[7]; John T. Mullen, MD[2]; Danielle Braun, PhD[3,8]; and Kevin S. Hughes, MD[2]

[1]Department of Surgery, Massachusetts General Hospital, Boston, MA, USA.
[2]Division of Surgical Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
[3]Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA, USA.
[4]Department of Breast Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Guangzhou, China.
[5]Department of General Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China.
[6]Computer Science & Artificial Intelligence, Massachusetts Institute of Technology, Boston, MA, USA.
[7]Department of Surgical Oncology, P. D Hinduja Hospital, Mumbai, India.
[8]Department of Biostatistics, Harvard University T.H. Chan School of Public Health, Boston, MA, USA.

**Correspondence**
Kevin S. Hughes, MD, FACS, Division of Surgical Oncology, Massachusetts General Hospital, 55 Fruit Street, Yawkey 7, Boston, MA 02114.
Email: kshughes@partners.org

## Abstract

**Background:** Pathogenic variants in germline cancer susceptibility genes can increase the risk of a large number of diseases. Our study aims to assess the disease spectrum of gastric cancer susceptibility genes and to develop a comprehensive resource of gene-disease associations for clinicians.

**Methods:** Twenty-seven potential germline gastric cancer susceptibility genes were identified from three review articles and from six commonly used genetic information resources. The diseases associated with each gene were evaluated via a semi-structured review of six genetic resources and an additional literature review using a natural language processing (NLP)-based procedure.

**Results:** Out of 27 candidate genes, 13 were identified as gastric cancer susceptibility genes (*APC, ATM, BMPR1A, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH-Biallelic, PALB2, SMAD4,* and *STK11*). A total of 145 gene-disease associations (with 45 unique diseases) were found to be associated with these 13 genes. Other gastrointestinal cancers were prominent among identified associations, with 11 of 13 gastric cancer susceptibility genes also associated with colorectal cancer, eight genes associated with pancreatic cancer, and seven genes associated with small intestine cancer.

**Conclusion:** Gastric cancer susceptibility genes are frequently associated with other diseases as well as gastric cancer, with potential implications for how carriers of these genes are screened and managed. Unfortunately, commonly used genetic resources provide heterogeneous information with regard to these genes and their associated diseases, highlighting the importance of developing guides for clinicians that integrate data across available resources and the medical literature.


**Keyword:** Gastric cancer; Cancer susceptibility gene; Disease spectrum; Germline; Cancer prevention.

## Introduction

Gastric cancer is a leading cause of cancer death, with over 1,000,000 new cases diagnosed worldwide per year (27,600 predicted cases in the United States in 2020) and over 780,000 deaths (11,010 in the United States) [1-3]. The importance of early diagnosis of gastric cancer is critical, as it is surgically curable if diagnosed early. Unfortunately, many gastric cancers are diagnosed at a later stage, and fewer than a third of patients with gastric cancer survive longer than five years [4].

At least some cases of gastric cancer are known to have a hereditary component. Approximately 10% of gastric cancers demonstrate aggregation within families, with an overall 1-3% thought to be truly hereditary [5]. For example, hereditary diffuse gastric cancer (HDGC) is an autosomal dominant gastric cancer syndrome primarily caused by pathogenic germline variants of the E-cadherin gene (*CDH1*). Current guidelines recommend individuals with *CDH1* pathogenic variants be evaluated for early prophylactic total gastrectomy to reduce gastric cancer risk [6, 7]. Additionally, individuals with pathogenic variants in *BMPR1A* and *SMAD4* [8-11], or those with Lynch syndrome [12-15] or Peutz-Jeghers syndrome [16, 17], are all considered to have an increased risk of gastric cancer.

While knowledge of a patient's genetic cancer risk affects screening and management for gastric cancer and can be a powerful tool in preventing gastric cancer death, there is increasing complexity in the genetics landscape. Clinicians may be faced with the task of interpreting genetic testing results of genes with which they have limited or no familiarity.

Even if a physician or provider is familiar with the increased risk of one particular type of cancer, they may not be fluent in the concomitant risk of other types of cancers resulting from the same pathogenic variant outside their specialty. For example, women with a *CDH1* pathogenic variant also have an increased risk of lobular breast cancer in addition to their increased risk of gastric cancer [18]. A number of resources exist to assist clinicians with patient counseling, surveillance, and medical management of patients with abnormal genetic test results, such as Clinical Genome Resource (ClinGen), the National Comprehensive Cancer Network (NCCN) [19-21], the Genetics Home Reference (now called MedlinePlus Genetics), Online Mendelian Inheritance in Man (OMIM), GeneCards, and Gene-NCBI [22-25]. Yet even when a physician/provider seeks out information regarding the meaning of a genetic test, there are inconsistencies between these sources regarding the clinical significance of genetic variants. The purpose of the current study is to review common clinical resources for germline genetic gastric cancer risk, to curate a list of genes identified to be associated with an increased risk of gastric cancer and to summarize the spectrum of diseases thought to be associated with these genes.

## Methods

The methods used to curate genes with a possible association with gastric cancer are similar to those used by Wang et al. [26] and are summarized in Figure 1. Briefly, a team consisting of three surgeons and four research fellows convened to review gene-disease associations for gastric cancer via a semi-structured process. First, potential gastric cancer susceptibility genes were curated. Then, these genes were scrutinized to identify associations with gastric cancer. Finally, those found to be associated with gastric cancer were studied to find what other diseases were associated with these genes. This process occurred from June 2019 to November 2020.



### Identifying possible gastric cancer susceptibility genes

Genes with possible gastric cancer association were identified from three recently published review articles (Identified by a PubMed search, not limited by year, using "gastric cancer," "stomach cancer," "genetic," and "mutation" as keywords) [27-29]. Twenty-seven genes (*APC, ATM, BMPR1A, BRCA1, BRCA2, CDH1, CHEK2, PALB2, CTNNA1, DOT1L, EPCAM, FBXO24, INSR, MAP3K6, MLH1, MSH2, MSH6, MSR1, MUTYH-Biallelic, MUTYH-Monoallelic, PMS2, PRSS1, PTEN, RAD51C, SMAD4, STK11,* and *TP53*) were identified as having any potential for a gastric cancer association.

### Verifying gene-gastric cancer associations based on six genetic resources and NLP

The associations between these genes and gastric cancer were then confirmed using all six genetic information resources that clinicians commonly refer to in understanding cancer risk for patients with particular genetic variants (ClinGen, NCCN, OMIM, Genetics

Home Reference, GeneCards, and Gene-NCBI) [19-25] as well as a natural language processing (NLP)-based literature review procedure [30, 31].

By resource, each gene was coded as being definitively associated with gastric cancer, possibly associated with gastric cancer, not associated with gastric cancer, or not mentioned. Because multiple resources were reviewed, each gene could have up to six gene-disease association codes. First, genes were reviewed with regard to their codes as generated by NCCN and ClinGen. As ClinGen is a database curated by genetic experts via a standardized framework, and NCCN uses the consensus of an expert panel in assessing genes, these two sources were considered as being as close to authoritative as available. If the association was verified by either source, it was considered accurate and was given a summary code of 'identified'. Genes specifically identified by ClinGen as not being associated with gastric cancer were given a summary code of ‘no association' with gastric cancer (NCCN does not explicitly state non-associations). Genes that were not confirmed or denied by ClinGen and/or NCCN but were identified as being definitively associated with gastric cancer in at least three of the remaining four resources, were also given a summary code of 'identified'. The remaining genes, including those with a discrepancy between NCCN and ClinGen or with fewer than three positive associations of the remaining four resources, were given an interim code of 'uncertain' and underwent additional scrutiny before the group made a final determination of the gene-disease association. Specifically, a formal literature review was completed using a semi-automated NLP-based procedure that searches and summarizes germline penetrance papers from PubMed. The sensitivity (99%) of this procedure in identifying cancer

penetrance studies and the workload reduction compared with the manual approach (84% reduction) have been validated [31]. Using standardized gene and disease search terms (Supplement 1), we ran a search query for each uncertain gene and identified gastric cancer-related penetrance studies. These papers were then reviewed. To establish a given gene's association with gastric cancer, we evaluated high-quality penetrance studies, which were selected based on their study design, patient population, number of pathogenic variant carriers, and ascertainment mechanism. If at least one high-quality penetrance study reported at least two-fold statistically significant increased risk, we considered a gene-cancer association as 'identified' [26]. Our database ultimately curated whether a gene was identified by each of the six resources plus the literature as having an association with gastric cancer.

Verifying the disease spectrum of gastric cancer susceptibility genes

Finally, using the same process described thoroughly in Wang et al. [26], the six resources were again reviewed for associations of other non-gastric cancer diseases with the list of gastric cancer susceptibility genes. To ensure the accuracy of the database two independent researchers reviewed each gene in all six resources for association with other diseases in a similar manner as described above. Any discrepancies between researchers were resolved through discussion at an in-person meeting until consensus was reached. The group consensus process was then repeated to review all gene-disease associations. Each association was then assigned a summary consensus code ("identified," "uncertain," or "no association") with regard to its association with a particular disease.

## Results

Gastric cancer susceptibility genes

After reviewing the six genetic resources and three recent review articles, a total of 27 possible gastric cancer susceptibility genes were curated (Table 1). Of these 27 genes, 13 genes were identified to be associated with gastric cancer (*APC, ATM, BMPR1A, CDH1, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH-Biallelic, PALB2, SMAD4,* and *STK11*). The remaining 14 genes were determined to have an uncertain association with gastric cancer.

Only *CDH1* was identified by all reviewed sources as being associated with gastric cancer (ClinGen, NCCN, OMIM, Genetics Home Reference, GeneCards, and Gene-NCBI). *APC* was identified by five of six resources (ClinGen, NCCN, OMIM, GeneCards, and Gene-NCBI), and *MLH1* and *MSH2* were identified as having a gastric cancer association by four of six resources (ClinGen, NCCN, Genetics Home Reference, and Gene-NCBI). The remainder of the verified gastric cancer association genes were identified by three or fewer resources. Nine of the 13 identified gastric cancer association genes were identified by both ClinGen and NCCN (*APC, BMPR1A, CDH1, EPCAM, MLH1, MSH2, MSH6, SMAD4,* and *STK11)*, one of the genes was identified by ClinGen but not NCCN (*MUTYH-Biallelic*), and one was identified by NCCN but not ClinGen (*PALB2*). Two of the identified genes were identified by neither ClinGen nor NCCN (*ATM* and *CHEK2*).

Disease spectrum of gastric cancer susceptibility genes

There were 190 potential gene-disease associations among the 13 identified gastric cancer susceptibility genes (Supplement 2). Including the 13 gene-gastric cancer associations, our group verified 145 gene-disease associations (with 45 unique diseases) (Table 2): 124 (85.5%) were noted by ClinGen and/or NCCN, and the other 21 were absent from both ClinGen and NCCN.

Among the 124 gene-disease associations that were noted by ClinGen and/or NCCN, 74 were identified by both ClinGen and NCCN, and the other 48 gene-disease associations were identified by either ClinGen or NCCN, but not both.

Among the 21 gene-disease associations that were absent from both ClinGen and NCCN, six were not identified by any of the six genetic resources but verified through NLP literature review alone (*CHEK2*-Thyroid cancer, Gastric cancer, Kidney cancer, and *MLH1/MSH2/MSH6*-Adrenocortical carcinoma). Among the remaining 15 gene-disease associations, four were verified through identification by three or more genetics resources (OMIM, Genetics Home Reference, GeneCards, and Gene-NCBI); the other 11 gene-disease associations were identified by one or two genetic resources but verified by NLP literature review.



Other gastrointestinal cancers associated with gastric cancer susceptibility genes

In addition to gastric cancer, other gastrointestinal cancers prominently associated with these 13 gastric cancer susceptibility genes were colorectal cancer, pancreatic cancer, and small intestine cancer. Eleven out of 13 gastric cancer susceptibility genes were

associated with colorectal cancer (*APC, ATM, BMPR1A, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH-Biallelic, SMAD4,* and *STK11*), eight genes were associated with pancreatic cancer (*APC, ATM, BMPR1A, EPCAM, MLH1, PALB2, SMAD4,* and *STK11*), and seven genes were associated with small intestine cancer (*APC, BMPR1A, EPCAM, MLH1, MSH2, SMAD4,* and *STK11*).

Several gastric cancer association genes are notable for being part of specific syndromes, such as *APC* for Familial Adenomatous Polyposis, *STK11* for Peutz-Jeghers, and multiple genes for Lynch syndrome (*MLH1, MSH2, MSH6,* and *EPCAM*). Of note, *PMS2*, another Lynch gene, was not judged to be definitively associated with gastric cancer.

## Discussion

Cancer prevention, screening, and early diagnosis are critical for improving gastric cancer prognosis, as most affected patients are diagnosed at a late stage. With the dramatic drop in DNA sequencing cost, cancer gene panel testing has become widely available and easy to access. Although different genes have different penetrance (i.e., the magnitude of cancer risk), knowing the gastric cancer susceptibility genes and their associated disease spectrum can still assist physicians in accurately identifying the high-risk patients and providing them with personalized prevention and treatment strategies, such as more frequent surveillance and screening for other cancers. This study used a semi-structured review process to assess the gene-gastric cancer association for 27 potential gastric cancer susceptibility genes and identified this association for 13 of these genes. The same process was used to identify the associated disease spectrum consisting of 145 gene-disease associations (with 45 unique diseases).

### Heterogeneity of commonly used genetic resources

One of the main findings of this work was confirmation of the significant heterogeneity of information that exists among six commonly used genetic resources, as previously noted by Wang et al. [26]. Only *CDH1* was recognized by all six sources as a gastric cancer susceptibility gene. For every other gene, a clinician could potentially fail to recognize an increased gastric cancer risk for a given patient, depending on which genetic resource was consulted. For example, *SMAD4* is identified as a gastric cancer association gene by both ClinGen and NCCN but not by OMIM, GeneCards, or Gene-NCBI, and *PALB2* was recognized only by NCCN and not by any of the other resources. Index of suspicion

may change how a provider interprets a patient's symptoms in the context of a known pathogenic variant, perhaps changing their threshold for referral for upper endoscopy for gastric cancer screening or risk-reducing gastrectomy. Delay in diagnosis or prevention may be devastating, as early gastric cancer is associated with significantly better survival [32]. Knowledge of the spectrum of diseases is also important as clinicians consider screening for non-gastric cancers in patients with known gastric cancer risk.

In the current work, the verification of the gene-disease associations was based almost entirely on examining the six genetics resources (Supplemented by NLP). It should be noted that these six genetic resources have variable updating frequencies: ClinGen and OMIM are updated regularly; Most NCCN guidelines are updated annually through an annual review process, while GHR, Gene cards, and Gene-NCBI are updated on an ongoing basis with each page updated separately. Of note, as of October 1, 2020, Genetics Home References ceased to exist as a stand-alone website, and most of its contents were transferred to MedlinePlus Genetics [33]. This variability in publication and review process among these six resources could be a reason behind the heterogenicity of the information available in each resource

As the standardized curation approach used by ClinGen is time-consuming and may lead to delay in reflecting the most recent findings and constantly changing evidence, and the gene-cancer associations listed on the NCCN guidelines may not be comprehensive, there is a significant need for a single resource that collects available information across multiple genetic resources as well as the medical literature. This reduces the chance that

a clinician may remain unaware of the potential consequences of a particular pathogenic variant in a relevant gene. This is especially important in the context of the increasing use of multi-gene panels, as clinicians may be faced with the task of interpreting genetic results for larger numbers of genes with which they may be unfamiliar [34].



Utility of NLP to assist in the literature review

Interestingly, several gene-disease associations were not identified in any of the six genetic resources but were identified by the NLP literature review. Given the exponential growth in medical literature, clinicians and researchers take more time and effort to extract relevant information. Developing semi-automated ways to search and parse the literature would allow individuals to more comprehensively identify and review useful information in less time. NLP is a subset of artificial intelligence (AI) that uses computational methods to extract meaning from natural human language. Given that scientific manuscripts often contain relevant information in narrative prose, NLP offers a promising way to review large numbers of scientific studies for relevance and glean critical data from those papers [35]. Our group has previously demonstrated the possibility of building an NLP-based medical abstract classifier to identify penetrance papers for gene-cancer associations [30, 31]. These NLP-based computational methods have permitted building an online risk calculator for cancer susceptibility genes called the All Syndromes Known to Man Evaluator™ (ASK2ME™), which has been recommended as a resource in recent clinical practice guidelines [36, 37].

Although the association between *CHEK2* and gastric cancer was not noted in any of these six genetic resources, we determined that *CHEK2* was associated with gastric cancer, kidney cancer, and thyroid cancer through the NLP-assisted literature review. Our NLP procedure identified a Polish, early-onset familial gastric cancer study that identified an association with *CHEK2* (OR = 2.1, p=0.01) [38]. This association was further characterized by a Danish population-based study in which the age- and sex-adjusted hazard ratio for gastric cancer in *CHEK2/1100delC* heterozygotes compared with noncarriers was 5.76 [39]. Within the same study, an association between *CHEK2* and kidney cancer was found (HR = 3.61) [39]. An association between *CHEK2* and thyroid cancer was also determined through an NLP-assisted literature review [40, 41]. The example of *CHEK2* and its associations with a number of cancers illustrates the benefit of using NLP for literature review, as relevant papers may be identified long before the primary literature is incorporated into intermittently curated genetic resources or guidelines.



Limitations

The conclusions regarding identified and uncertain gene-disease associations are only as reliable as the data sources and the procedures used to adjudicate conflicting gene-disease associations. With regard to data sources, we in large part relied on the curation processes of the six genetic resources to identify an association between a particular gene and a disease. Therefore, any limitation of each resource's curation process will affect the reliability of our results. We tried to mitigate this weakness by a consensus review of all six genetic resources plus an NLP-aided literature review. Additionally, all

these genetic resources are updated periodically, and it is possible that future research efforts will clarify gastric cancer susceptibility genes and disease associations that we concluded were uncertain. Therefore, this study represents a snapshot of current knowledge and understanding of gastric cancer susceptibility genes, rather than a definitive conclusion.

This second data limitation underscores the importance of having a continually updated database of gene-disease associations, as the literature is rapidly expanding beyond the capability of any given clinician to track genetic risk factors for a given disease. Our review process required human review of the six genetic resources to code whether each resource reported an association between a given gene and a disease. Two individual researchers independently coded each gene-disease association, and their separate coding results were brought to a group for consensus review and confirmation. Additionally, we acknowledge that the semi-structured process was internally developed. However, there currently is no "gold standard" when reviewing gene-disease associations across multiple resources or the broader medical literature, and our procedure bears a number of features that we believe increase its rigor and reliability, i.e., two coders were used to independently generate the database, a group of researchers reviewed and discussed each gene-disease association, and an NLP-assisted literature review was incorporated to increase the comprehensiveness of the review.

The third limitation is the clinical actionability and relevance of the results. With regard to gastric cancer, clinicians must consider whether a particular pathogenic variant places a

patient at sufficient risk to warrant changes in management, such as referral for a screening upper endoscopy or a prophylactic total gastrectomy. In the case of the *CDH1* pathogenic variant, the high penetrance of the gene resulting in high lifetime rates of gastric cancer and the inadequacy of endoscopy in identifying this disease early have led to the recommendation for prophylactic total gastrectomy [7]. Yet other gene-disease associations are likely far less strong, with significantly lower penetrance and risk of gastric cancer or with causing a type of gastric cancer that is more detectable by frequent endoscopy. For the identified gastric cancer susceptibility genes, we are unable to report penetrance, and therefore the optimal impact on clinical decision making is yet to be determined. NCCN guidelines for gastric cancer acknowledge that even for syndromes with a known estimated increase in gastric cancer, such as Familial Adenomatous Polyposis, there is no clear evidence on which to base the recommendation of a first screening upper endoscopy between the ages of 25 and 30. Determining whether the degree of gastric cancer risk conferred by each of our 13 identified gastric cancer susceptibility genes should influence clinical care is beyond the scope of this current manuscript. Realistically, actively monitoring carriers of gastric cancer susceptibility genes may not be feasible due to a number of factors, including cost-effectiveness and patient preference (i.e., patient anxiety from frequent medical testing), and guidelines need to be updated as more data becomes available.

Future work

Identifying gastric cancer susceptibility genes and their associated diseases is an important first step in identifying individuals for whom screening and monitoring may prove to be beneficial, as gastric cancer is most likely curable when diagnosed at an early stage. However, prior to influencing clinical guidelines, more work to understand the penetrance and clinical impact of these 13 genes must be undertaken. Using an NLP algorithm to identify and extract relevant data from the medical literature may assist researchers in incorporating all available data to more accurately estimate the effects of genetic contributors to gastric cancer. We aim to collect and present data regarding gastric cancer and other malignancies in a single resource, ASK2ME™, to help scientists and clinicians understand and interpret pathogenic variants' clinical consequences.

## Conclusion

We developed a semi-structured review process, incorporating the review of six genetic resources as well as an NLP-based literature review, to identify and collate gastric cancer susceptibility genes and their associated diseases into a single publication. In this process, we demonstrated significant information heterogeneity of prominent genetic resources. Variation in genetic risk identification could result in provider confusion, missed opportunities for genetic testing and counseling, and suboptimal clinical decision-making. Notably, we confirmed that gastric cancer susceptibility genes are frequently associated with many other cancers as well, with potential implications for how these patients are screened and managed. Future work to accurately estimate the penetrance of these gene-cancer associations and calculate disease risk will further inform clinical management and decision-making for individuals affected by gastric cancer-associated genes. Constantly updated databases such as ASK2ME™ may serve as important resources of curated information so that individual clinicians do not have to independently review an ever-increasing quantity of medical literature.

**Acknowledgements**

The authors acknowledge Ann S. Adams (Department of Surgery, Massachusetts General Hospital) for editorial and writing assistance.

**Declarations**

**Funding**: This study received no specific funding.

**Conflicts of interest/Competing interests**: Kevin S. Hughes receives Honoraria from Hologic (Surgical implant for radiation planning with breast conservation and wire free breast biopsy), and Myriad Genetics and has a financial interest in CRA Health (Formerly Hughes RiskApps). CRA Health develops risk assessment models/software with a particular focus on breast cancer and colorectal cancer. Dr. Hughes is a founder and owns equity in the company. Dr. Hughes is the Co-Creator of Ask2Me.Org which is freely available for clinical use and is licensed for commercial use by the Dana Farber Cancer Institute and the MGH. Dr. Hughes's interests in CRA Health and Ask2Me.Org were reviewed and are managed by Massachusetts General Hospital and Partners Health Care in accordance with their conflict-of-interest policies. Dr. Braun co-leads the BayesMendel laboratory, which licenses software for the computation of risk prediction models. She does not derive any personal income from these licenses. All revenues are assigned to the lab for software maintenance and upgrades. The other authors declare that they have no conflict of interest.

**Ethics approval**: We used public databases with no patient data, and ethical committee was waived.

**Availability of data and material**: All data used within this research are available from the corresponding author upon reasonable request.

## References


1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. International Agency for Research on Cancer. Cancer today. [Available from: http://gco.iarc.fr/today/online-analysis-pie?v=2018&mode=cancer&mode_population=continents&population=900&populations=900&key=total&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&nb_items=7&group_cancer=1&include_nmsc=1&include_nmsc_other=1&half_pie=0&donut=0&population_group_globocan_id= ]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.
4. Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Stomach Cancer. [Available from: https://seer.cancer.gov/statfacts/html/stomach.html.]
5. Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015;16(2):e60-70.
6. van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, et al. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J Med Genet. 2015;52(6):361-74.
7. National Comprehensive Cancer Network. Gastric Cancer. [Available from: https://www.nccn.org/professionals/physician_gls/pdf/gastric.pdf]

8. Li J, Woods SL, Healey S, Beesley J, Chen X, Lee JS, et al. Point mutations in exon 1B of APC reveal gastric adenocarcinoma and proximal polyposis of the stomach as a familial adenomatous polyposis variant. Am J Hum Genet. 2016;98(5):830-42.
9. Worthley DL, Phillips KD, Wayte N, Schrader KA, Healey S, Kaurah P, et al. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut. 2012;61(5):774-9.
10. Tedaldi G, Pirini F, Tebaldi M, Zampiga V, Cangini I, Danesi R, et al. Multigene panel testing increases the number of loci associated with gastric cancer predisposition. Cancers (Basel) 2019;11(9):1340.
11. Chun N, Ford JM. Genetic testing by cancer site: stomach. Cancer J. 2012;18(4):355-63.
12. Bonadona V, Bonaïti B, Olschwang S, Grandjouan S, Huiart L, Longy M, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA. 2011;305(22):2304-10.
13. Møller P, Seppälä TT, Bernstein I, Holinski-Feder E, Sala P, Gareth Evans D, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. Gut. 2018;67(7):1306-16.
14. Engel C, Loeffler M, Steinke V, Rahner N, Holinski-Feder E, Dietmaier W, et al. Risks of less common cancers in proven mutation carriers with lynch syndrome. J Clin Oncol. 2012;30(35):4409-15.

15. Capelle LG, Van Grieken NCT, Lingsma HF, Steyerberg EW, Klokman WJ, Bruno MJ, et al. Risk and epidemiological time trends of gastric cancer in Lynch syndrome carriers in the Netherlands. Gastroenterology 2010;138(2):487-92.
16. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJP, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res. 2006;12(10):3209-15.
17. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 2000;119(6):1447-53.
18. Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, et al. Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond. JAMA Oncol 2015;1(1):23-32.
19. ClinGen. Gene Validity Curations. [Available from: https://search.clinicalgenome.org/kb/gene-validity]
20. NCCN Clinical Practice Guidelines in Oncology. [Available from: https://www.nccn.org/professionals/physician_gls/default.aspx]
21. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen--the Clinical Genome Resource. N Engl J Med. 2015;372(23):2235-42.
22. Genetics Home Reference. [Available from: https://ghr.nlm.nih.gov/]
23. Home - OMIM - NCBI. [Available from: https://www.ncbi.nlm.nih.gov/omim]
24. GeneCards Human Gene Database. [Available from: https://www.genecards.org
25. Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., eds. Gene-NCBI®. Seattle (WA): University of Washington, Seattle, 2010

26. Wang J, Singh P, Yin K, Zhou J, Bao Y, Wu M, et al. Disease spectrum of breast cancer susceptibility genes. medRxiv doi: 10.1101/2020.08.11.2017200
27. Sahasrabudhe R, Lott P, Bohorquez M, Toal T, Estrada AP, Suarez JJ, et al. Germline mutations in PALB2, BRCA1, and RAD51C, which regulate DNA recombination repair, in patients with gastric cancer. Gastroenterology 2017;152(5):983-6.e6
28. Petrovchich I, Ford JM. Genetic predisposition to gastric cancer. Semin Oncol. 2016;43(5):554-9.
29. Slavin TP, Weitzel JN, Neuhausen SL, Schrader KA, Oliveira C, Karam R. Genetics of gastric cancer: what do we know about the genetic risks? Transl Gastroenterol Hepatol. 2019;4:55.
30. Bao Y, Deng Z, Wang Y, Kim H, Armengol VD, Acevedo F, et al. Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes. JCO Clin Cancer Inform. 2019;3:1-9.
31. Deng Z, Yin K, Bao Y, Armengol VD, Wang C, Tiwari A, et al. Validation of a semi-automated natural language processing-based procedure for meta-analysis of cancer susceptibility gene penetrance. JCO Clin Cancer Inform. 2019;3:1-9.
32. American Cancer Society. Stomach Cancer Survival Rates. [Available from: https://www.cancer.org/cancer/stomach-cancer/detection-diagnosis-staging/survival-rates.html#references ]
33. MedlinePlus - Genetics [Available from: https://medlineplus.gov/genetics/ ]
34. Plichta JK, Griffin M, Thakuria J, Hughes KS. What's New in Genetic Testing for Cancer Susceptibility? Oncology (Williston Park). 2016;30(9):787-99.

35. Hughes KS, Zhou J, Bao Y, Singh P, Wang J, Yin K. Natural language processing to facilitate breast cancer research and management. Breast J. 2020;26(1):92-9.
36. Braun D, Yang J, Griffin M, Parmigiani G, Hughes KS. A clinical decision support tool to predict cancer risk for commonly tested cancer-related germline mutations. J Genet Couns. 2018;27(5):1187-99.
37. Manahan ER, Kuerer HM, Sebastian M, Hughes KS, Boughey JC, Euhus DM, et al. Consensus guidelines on genetic testing for hereditary breast cancer from the American Society of Breast Surgeons. Ann Surg Oncol. 2019;26(10):3025-31.
38. Teodorczyk U, Cybulski C, Wokołorczyk D, Jakubowska A, Starzyńska T, Lawniczak M, et al. The risk of gastric cancer in carriers of CHEK2 mutations. Fam Cancer. 2013;12(3):473-8.
39. Näslund-Koch C, Nordestgaard BG, Bojesen SE. Increased Risk for Other Cancers in Addition to Breast Cancer for CHEK2*1100delC Heterozygotes estimated from the Copenhagen General Population Study. J Clin Oncol. 2016;34(11):1208-16.
40. Kaczmarek-Ryś M, Ziemnicka K, Hryhorowicz ST, Górczak K, Hoppe-Gołębiewska J, Skrzypczak-Zielińska M, et al. The c.470 T > C CHEK2 missense variant increases the risk of differentiated thyroid carcinoma in the Great Poland population. Hered Cancer Clin Pract. 2015;13(1):8.
41. Siołek M, Cybulski C, Gąsior-Perczak D, Kowalik A, Kozak-Klonowska B, Kowalska A, et al. CHEK2 mutations and the risk of papillary thyroid cancer. Int J Cancer. 2015;137(3):548-52.

## Figure legend

**Figure 1. Flow chart for identifying and evaluating gene-disease association.** The number '1' indicates that the gene was associated with gastric cancer in the resource. The number '0' indicates that the gene's association with gastric cancer was refuted in the resource. The number '9' indicates that the gene's association with gastric cancer was unclear in the resource.

*Uncertain association indicates that the gene's association with gastric cancer is unclear and it may or may not be associated with gastric cancer, further studies are required to refute or accept the association.

**Table 1. Association for 27 candidate genes with gastric cancer in six genetic resources.** The '+' sign indicates that the gene was associated with the disease in the resource. Blank space indicates that the association was not found in the resource.

| Gene | Genetic Resources | | | | | | |
|---|---|---|---|---|---|---|---|
| | ClinGen | NCCN | GHR | OMIM | GeneCards | Gene-NCBI | Co |
| ***APC*** | + | + | | + | + | + | Ide |
| ***ATM*** | | | + | | | | Ide |
| ***BMPR1A*** | + | + | | | | + | Ide |
| ***BRCA1*** | | | + | | | | Un |
| ***BRCA2*** | | | | | | | Un |
| ***CDH1*** | + | + | + | + | + | + | Ide |
| ***CHEK2*** | | | | | | | Ide |
| ***CTNNA1*** | | | + | | + | | Un |
| ***DOT1L*** | | | | | | | Un |
| ***EPCAM*** | + | + | | | | + | Ide |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ***FBXO24*** | | | | | | | **Un** |
| ***INSR*** | | | | | | | **Un** |
| ***MAP3K6*** | | | | | **+** | | **Un** |
| ***MLH1*** | **+** | **+** | **+** | | | **+** | **Ide** |
| ***MSH2*** | **+** | **+** | **+** | | | **+** | **Ide** |
| ***MSH6*** | **+** | **+** | **+** | | | | **Ide** |
| ***MSR1*** | | | | | | | **Un** |
| ***MUTYH-Biallelic*** | **+** | | | | | | **Ide** |
| ***MUTYH-Monoallelic*** | | | | | | | **Un** |
| ***PALB2*** | | **+** | | | | | **Ide** |
| ***PMS2*** | | | **+** | | | **+** | **Un** |
| ***PRSS1*** | | | | | **+** | | **Un** |
| ***PTEN*** | **+** | | | | | | **Un** |
| ***RAD51c*** | | | | | | | **Un** |
| ***SMAD4*** | **+** | **+** | | | | **+** | **Ide** |
| ***STK11*** | **+** | **+** | | **+** | | | **Ide** |
| ***TP53*** | | | | **+** | | | **Un** |

Abbreviation: ClinGen, Clinical Genome Resource; NCCN, The National Comprehensive Cancer Network; OMIM, Online Mendelian Inheritance in Man; GHR, Genetics Home Refence

**Table 2. Diseases associated with 13 gastric cancer susceptibility genes**. *These gene-disease associations were identified using all 6 genetic resources as well as NLP literature review.

| Gastric Cancer Susceptibility Gene | Disease spectrum | | |
|---|---|---|---|
| | **Malignant** | **Benign** | **Borderline** |
| ***APC*** | Brain cancer, Colorectal cancer, Duodenal cancer, Gastric cancer, Hepatobiliary cancer, Pancreatic cancer, Small intestine cancer, Thyroid cancer | CHRPE, Skin disorders, Tooth disorders | Adrenal neoplasm, Bone neoplasm, Soft tissue neoplasm |
| ***ATM*** | Breast cancer, colorectal cancer, Gastric cancer, Pancreatic cancer, Prostate cancer | | |
| ***BMPR1A*** | Colorectal cancer, Duodenal cancer, Gastric cancer, Pancreatic cancer, Small intestine cancer | | |
| ***CDH1*** | reast cancer, Gastric cancer | Eye disorders, Facial dysmorphism, Gastrointestinal disorders, Hair disorder, Hand disorder, Nail disorder, Nose disorder, Orofacial cleft disorder, Tooth disorder, Thyroid disorder | |
| ***CHEK2*** | Breast cancer, Colorectal cancer, Gastric cancer, Kidney cancer, Prostate cancer, Sarcoma, Thyroid cancer | | |
| ***EPCAM*** | Bladder cancer, Brain cancer, Colorectal cancer, Endometrial cancer, Gastric cancer, Hepatobiliary cancer, Kidney cancer, Ovarian cancer, Pancreatic cancer, Prostate cancer, Sebaceous cancer, Small intestine cancer, Urinary tract cancer, Ureteral cancer | Skin disorder | |

| | | | |
|---|---|---|---|
| ***MLH1*** | Adrenocortical carcinoma, Bladder cancer, Brain cancer, Colorectal cancer, Endometrial cancer, Gastric cancer, Hepatobiliary cancer, Ovarian cancer, Pancreatic cancer, Prostate cancer, Sebaceous cancer, Small intestine cancer, Urinary tract cancer, Ureteral cancer | Skin disorder | |
| ***MSH2*** | Adrenocortical carcinoma, Bladder cancer, Brain cancer, Colorectal cancer, Endometrial cancer, Gastric cancer, Hepatobiliary cancer, Kidney cancer, Ovarian cancer, Prostate cancer, Sebaceous cancer, Small intestine cancer, Urinary tract cancer, Ureteral cancer | Skin disorder | |
| ***MSH6*** | Adrenocortical carcinoma, Bladder cancer, Brain cancer, Colorectal cancer, Endometrial cancer, Gastric cancer, Hepatobiliary cancer, Kidney cancer, Ovarian cancer, Prostate cancer, Sebaceous cancer, Urinary tract cancer, | Benign skin disorder | |
| ***MUTYH-Biallelic*** | Colorectal cancer, Duodenal cancer, Gastric cancer | CHRPE, Skin disorder | |
| ***PALB2*** | Breast Cancer, Gastric Cancer, Ovarian Cancer, Pancreatic Cancer, Prostate Cancer | | |
| ***SMAD4*** | Colorectal cancer, Duodenal cancer, Gastric cancer, Pancreatic cancer, Small intestine cancer | Bone disorder, Cardiovascular disease, Lung disease | |
| ***STK11*** | Breast cancer, Cervical cancer, colorectal cancer, Endometrial cancer, Gastric cancer, Hepatobiliary cancer, Lung cancer, Pancreatic cancer, Small intestine cancer | Skin disorder, Gastrointestinal hamartomatous polyps | Testicular Neoplasm, Ovarian Neoplasm |

Abbreviation: CHRPE, congenital hypertrophy of the retinal pigment epithelium.