Genetic Variations and Gastric Cancer

Chen-Yang Yu  Hao-Yan Chen
Division of Gastroenterology and Hepatology, Renji Hospital, Shanghai Jiaotong University
School of Medicine, Shanghai Institute of Digestive Disease, Shanghai, China

Key Words
Expression quantitative trait loci · Gastric cancer · Genetic variations · Genome-wide association study · Single nucleotide polymorphism

Abstract

Background: Gastric cancer (GC) has an apparent hereditary component. However, in a large fraction of gastric cases, no known genetic syndrome or family history can be identified, suggesting the presence of ‘missing heritability’ in GC etiology. Genome-wide association studies (GWAS) and traditional candidate gene studies have both led to the identification of multiple replicable common genetic variants associated with GC risk. Summary: We summarize the genetic variants associated with GC risk identified up to date. Achievements derived from translational cancer research including the following aspects: (a) contribution to the our understanding of gastric tumorigenesis, (b) guidance to individualized treatment and (c) prediction of patient prognosis. We also prospect future research direction such as post-GWAS analyses and rare variants studies. Key Message: Many genetic variants were found through GWAS or candidate gene studies, and interpreting their underlying mechanisms will help us translate risk profiles generated from these variations into use in the clinical setting for targeted screening and treatment. Practical Implications: Investigation of the potential use of genetic variations as prognostic and predictive markers is a developing field. Many people could benefit from a better understanding of genetic polymorphisms to potentially identify a priori individuals who might have the best chance of survival and therefore derive most clinical benefit from treatment. Outcomes of particular scientific interest for molecular epidemiologic studies should include overall survival, recurrence- and progression-free survival, response to treatment, and early and late toxicities stemming from chemotherapy and radiation.

Hao-Yan Chen, MD, PhD
Division of Gastroenterology and Hepatology, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai Institute of Digestive Disease
Shanghai 200001 (China)
E-Mail haoyanchen@shsmu.edu.cn

© 2015 S. Karger AG, Basel
Introduction

Gastric cancer (GC) is the fourth most common cancer and the second leading cause of cancer deaths worldwide, although its incidence and mortality have been appreciably declining for several decades [1, 2]. Germline variations of DNA sequence are believed to represent a key aspect of predisposition to most complex traits, such as cancer [3, 4]. Single nucleotide polymorphisms (SNPs) are naturally existing genetic variations that occur with variable frequency within different ethnic populations. Particular SNPs can alter the gene expression profile and affect gene function, which leads to increased risk of susceptibility to a range of diseases, including cancer. Many examples exist of polymorphic genes that increase the susceptibility to GC [5–8]. This relationship was initially investigated within chosen candidate genes in a hypothesis-driven manner that was based on disease pathophysiology. Now, with the advent of improving technologies, genome-wide association studies (GWAS) and high-throughput genetic analysis, information is available on multiple previously unidentified SNPs within a large number of genes. These new approaches mean that the SNPs can be simultaneously assessed, offering new insights into the pathogenesis of GC. The frequency of particular genetic polymorphisms, as well as which polymorphisms are most common, varies in accordance with ethnicity, and together with exposure to environmental risk factors can promote the development of malignant disease [9].

Discovery of Genetic Variants

Candidate gene research interrogates a limited number of variants with known mechanistic importance; therefore, it is not surprising that the majority of candidate gene studies have focused on known cancer-related genes. On the other hand, GWAS platforms are designed to account for tagged SNPs with significant linkage disequilibrium among multiple variants, rather than focusing on functional polymorphisms with known mechanistic importance.

For this reason, GWAS have generally found different variants in association with GC risk and outcome, and these same variants typically have a rather unclear mechanistic relationship with known disease etiology. Additional factors also confound studies of genetic markers, given that genetic penetrance is often low in disease etiology and progression, and numerous other multifactorial complexities are taken for granted. GC is a heterogeneous disease with a multigenic origin, so these complexities are likely to confound studies of GC in particular. Both candidate gene and GWAS approaches also have their own statistical limitations. For instance, candidate gene studies typically have a high rate of false positives and overestimated effect sizes, while genome-wide approaches typically have a high rate of false negatives due to multiple testing. It is therefore challenging to ascertain which type of study is more valid, and it is more likely that the answer lies in between – that we need to be focused on ascertaining the true importance of candidate gene findings while ascertaining the mechanism of properly conducted GWAS.

Variants from GWAS

With the arrival of high-throughput genotyping techniques for whole genomes, many germline variants have been discovered that are significantly associated with GC susceptibility [6–8, 10, 11]. While the use of GWAS, next-generation sequencing, whole exome sequencing and RNA sequencing has allowed for the comprehensive analysis of GC genomes, it has also given an indication of the complexity and heterogeneous nature of GC. The use of these new tools has led to the identification of multiple chromosome loci, gastric-, metabolism- and inflammation-specific genes, somatic alterations of GC genomes, fusion transcripts and gene fusions.
One common result of GWAS has been the identification of 8q24, and prostate stem cell antigen (PSCA) has been identified in this region of biological significance. Recent studies have identified loci at 1q22, 10q23, 3q13 and 5p13 that all demonstrate gastric-specific genes associated with gastric development and carcinogenesis. GWAS have identified over five variants or loci shown to be significantly associated with the risk of GC. Unfortunately due to the modest effect by these markers, their use in GC risk assessment is limited. Also limiting the use of GWAS and other high-throughput technology is the sample size of most current studies due to cost and sample acquisition, the complexity of the GC genome and the ethnic and racial group differences in results. Interpreting the results of the enormous amount of GWAS data and applying the findings to the clinic remains a challenge as most of the identified loci currently lack biological significance. This problem will be overcome in the future as better bioinformatic approaches become available to help define the biology behind these associations.

**Variants from Candidate Gene Studies**

Candidate gene studies are studies whereby investigators assume that a specific gene, and typically a specific functional gene polymorphism that results in alterations in protein function or quantity, may play a role in disease pathophysiology. So far, numerous candidate gene studies have been conducted in the area of GC research.

Many investigators have reported associations between SNPs in genes that regulate the host’s inflammatory response and GC. The inflammatory response-related genes that have been most frequently studied in relation to GC are interleukin (IL) genes IL1B, IL1RN, IL8, IL10, IL17 and tumor necrosis factor alpha (TNF-α), coding for the proteins IL-1b, IL-1ra, IL-8, IL-10, IL-17 and TNF-α, respectively. These cytokines are important mediators in gastric physiology and pathophysiology and could play important roles in the etiology of GC. In 2000, El-Omar et al. [5] reported that interleukin-1 gene cluster polymorphisms are associated with an increased risk of both hypochlorhydria induced by *Helicobacter pylori* and GC. An overwhelming majority of studies characterize the T allele of the IL1B-511 polymorphism as frequent among Caucasian individuals with non-cardia GC, preferably for the intestinal subtype of this cancer [12–15].

In recent years, studies have confirmed that SNPs in the promoter region of the CDH1 gene influence its transcriptional activity and alter the expression of E-cadherin. The most widely studied polymorphism is CDH1 C-160A, where the A allele decreases transcription efficiency of the CDH1 gene and may increase susceptibility to cancer development in some populations. Recently, a considerable number of studies have been conducted to investigate the associations between the CDH1 C-160A polymorphism and susceptibility for GC [16–19]. However, a recent meta-analysis failed to confirm the association between the CDH1 C-160A polymorphism and risk of GC [20].

**Interpretation**

GWAS have identified more than 200 mostly new common low-penetrance susceptibility loci for cancers. The predicted risk associated with each locus is generally modest and so, presumably, are the functional effects of individual genetic variants conferring disease susceptibility. Perhaps the greatest challenge in the ‘post-GWAS’ era is to understand the functional consequences of these loci. Biological insights can then be translated into clinical benefits, including reliable biomarkers and effective strategies for screening and disease prevention [21–23].
Expression Quantitative Trait Loci (eQTL)

The core hypothesis behind eQTL analysis is that polymorphic sites in the genome, such as SNPs, could have a tangible effect on gene regulation by altering the coding or promoter sequences of genes, their splicing junctions or other regulatory elements. All of these regions affect the rate at which genes are transcribed, which isoforms are preferentially expressed and how stable the final messenger RNA product is. Thus, SNPs suspected of affecting gene expression (eSNPs) can be tested with associative statistics. Two types of eQTL can generally be differentiated from one another: cis and trans. Polymorphic sites within chromosomal proximity of a gene affecting its expression are considered to be cis regulators. In contrast, elements elsewhere on the genome are thought to be acting in trans. The distinction between cis and trans, however, is not always well-defined. With the exception of a few straightforward possibilities when the polymorphic site falls within the start and end coordinates of a gene, or when it is on a different chromosome altogether, labeling a locus as either cis or trans is determined by a rigid distance from a gene measured in base pairs. Commonly used intervals around genes to define cis action have ranged from a few tens of thousands of bases to millions [24, 25].

Genetic variants located within cancer-related genes may affect their expressive levels. SNP rs869736 located at position -309 of the PTPRCAP promoter was associated with susceptibility to diffuse-type GC in the current case-control study. Furthermore, the cancer-risk, minor-allele T of rs869736 increased both promoter activity and specific nuclear protein-binding affinity than the non-risk, major-allele G in luciferase reporter and electrophoretic mobility shift assays, respectively [26]. The DNMT3A -448A>G polymorphism is also a functional SNP and contributes to its genetic susceptibility to GC. In a promoter assay, carriage of the -448A allele showed a significantly higher promoter activity compared with the -448G allele (p < 0.001) [27]. Genetic variants could also affect expressive levels of non-coding RNAs [28–30].

Another group of eSNPs located within 3′-UTR regions affect expressive levels of cancer-related genes through impact on binding ability between miRNA and its target genes. Wang et al. [31] found that a common C>G polymorphism (rs4143815) was significantly different between cases and controls (p = 1.32 × 10^{-8}). Luciferase reporter assay indicated that this SNP might be responsible for aberrant B7-H1 protein expression in GC by disrupting the interaction between miR-570 and B7-H1 mRNA. Song et al. [32] found a significantly increased risk of GC associated with the SCRN1 rs6976789 C>T polymorphism (adjusted odds ratio = 1.25, 95% confidence interval = 1.02–1.53; CT/TT vs. CC). Luciferase report assay showed that the rs6976789 variant T allele influenced the binding ability of miR-148a. Similar research have been done in GC [33–35].

Clinical Implications of Variants

Association of Variants with Treatment Strategy. Although the majority of genetic studies to date are aimed at identifying predisposition variants, other genetic studies have also been conducted in GC treatment cohorts with the aim of identifying germline variants that could predict treatment outcomes or toxicities. Multiple options are currently available for the treatment of GC, from surgery and radiotherapy to systemic therapy with cytotoxic agents and targeted therapies. Toxicities and responses to each of these treatment modalities can vary, the mitigating factors of which remain unclear, although genetic variants have been proposed and implicated. In 2006, Lu et al. [36] found that the presence of the thymidylate synthase (TS) 3′-UTR 6 bp nucleotide fragment can be correlated with the sensitivity of GC to 5-FU-based chemotherapy (CTx), and that the TS 3′-UTR polymorphism profile can be used to guide the choice of 5-FU-based CTx in advanced GC. Ott et al. [37] evaluated DNA polymorphisms in the TS and 5,10-methylene tetrahydrofolate reductase (MTHFR) genes for an association with response and survival in locally advanced GC treated with 5-FU-based preoperative CTx. Comparing survival between R0 patients with and without CTx in the respective TS
genotype groups of the tandem repeat polymorphism, a significant survival benefit for patients with CTx was found for the 2rpt/2rpt (n = 49; p = 0.002) and 2rpt/3rpt genotypes (n = 99; p = 0.004), but not for the 3rpt/3rpt genotype (n = 57; p = 0.93). Patient survival after CTx was associated with the TS tandem repeat polymorphism. Cytotoxicity analysis done by Ha et al. in 2011 [38] showed that most of the MKN1 and SNU638 clones transfected with the G allele of deoxyribonuclease II beta rs3738573 were more sensitive to docetaxel than those with the C allele (p ≤ 0.001–0.029) and most of the AGS and SNU638 clones transfected with the T allele of 5-hydroxytryptamine receptor IE rs3828741 were more sensitive to paclitaxel than those with the C allele.

Association of Variants with Prognosis. Some genetic variants are associated with patients’ prognosis of GC. García-González et al. [39] identified the rs2294008T variant as a prognosis factor associated with worse overall survival in patients with diffuse-type GC (hazard ratio = 1.85; 95% confidence interval = 1.12–3.06). PSCA rs2294008 polymorphism is involved in the prognosis of the diffuse type of GC in Caucasians. Wang et al. [40] performed a genetic association study to analyze the correlation between the five tagged SNPs rs4135385, rs1798808, rs1880481, rs11564465 and rs2293303 of CTNNB1 and GC risk and survival. Only the SNP rs4135385 AG/AA genotypes were significantly associated with a favorable GC survival compared with the GG genotype (adjusted hazard ratio = 0.80; 95% confidence interval = 0.66–0.97), and the association was more prominent among patients with non-cardia GC than those with cardia GC (log-rank p = 0.007 for non-cardia GC and 0.417 for cardia GC). Polymorphisms in genes such as SCRNI, MTHFR, MAP3K1, TLR-9 and pri-let-7a-2 also have been demonstrated to be useful prognostic biomarkers for GC patients [32, 41–46].

Functional Investigation

Increased investment in post-GWAS functional characterization of risk loci has now been advocated across diseases and for cardiovascular disease and diabetes. For cancer biology, the complex interplay between genetics and the environment in many cancers poses a particularly exciting challenge for post-GWAS research [21]. As for genetic variants in GC, only a small portion of SNPs have been functionally interpreted.

Functional analysis revealed that the PSCA protein is expressed in areas of the glandular crypts along with stem cell progenitors, which is the proposed initiator cell for this particular histological subtype of GC. In GC tissue specimens, PSCA was downregulated at both the gene and protein level. To unravel the biological consequence of this finding, in vitro transfection studies were undertaken that revealed a role for PSCA in the inhibition of epithelial cell proliferation. This unexpected molecular association has subsequently been validated in several case-control studies in Asian populations, albeit with variable frequency and with the risk of both intestinal and diffuse histological subtypes [6].

An assessment of MUC1 protein expression in GC specimens in accordance with genotype found that the A allele was associated with reduced protein expression in GC tissue. This polymorphism was found to affect the function of the gene promoter in functional reporter assays, with deletion of the rs4072037 loci leading to reduced transcriptional activity. In addition, presence of the G allele confers higher activity than the A allele. This polymorphism can regulate splicing variation. Carriage of the A allele results in deletion of a nine amino acid segment in exon 2 with downstream protein modification [8].

Perspective

It is always tempting to speculate on the genetic mechanisms and molecular pathways by which GWAS loci affect the risk of GC. Some loci are near genes with known functions, indi-
cating these may be good candidates for being the functionally relevant gene. Others lie in ‘gene deserts’ with no nearby genes. Despite much speculation, the functional effects and target genes at most of the confirmed loci have yet to be elucidated, while those that have been examined in detail have often generated surprising results. Post-GWAS analyses include detailed genetic epidemiological dissection of the associated locus, bio-informatic prediction of functionality and in vitro and in vivo experimental verification of the molecular mechanisms for the causal variants and their target genes [47]. The initial epidemiological studies require dense SNP genotyping in large sample sizes, first to analyze the effects of less common candidate variants (minor frequency 1–5%), and second to separate neighboring genetic variants that are often correlated. Chakravarti et al. [48] suggested that to formally establish causality of a given variant it is necessary to demonstrate in human cells/tissues or animal models that recreating the risk variants would generate analogous phenotypes in the model system.

There are good reasons to expect that common genetic variants do not explain all of the inherited risk of GC, not the least of these being the relatively low proportion of familial relative risk that common cancer SNPs currently explain. One promising source of the unexplained risk are rare, low-penetration genetic variants, a class that ranges from low-frequency polymorphisms (allele frequency <5%) through subpolymorphic variants (frequency 0.1–1.0%) to very-low-frequency or ‘private’ variants with frequencies of 0.1% or less. Several rare variants implicating tumorigenesis have been investigated [49–51]; however, there are almost no detailed studies about rare variants related to GC.

Because of cost considerations, current rare variant studies largely focus on exome regions. We expect that the focus will gradually extend as the cost of whole genome sequencing decreases. Challenges for whole-genome rare variant analysis include limited available information for prioritizing and annotating most likely functional variants, which is important for grouping variants for multi-marker tests and interpreting results. Progress in annotating the functional consequences of non-exome variants will facilitate future genome-wide sequencing-based association studies.

**Disclosure Statement**

No potential conflicts of interest are disclosed.

**References**


