Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma

1Cancer and Stem Cell Biology Program, Duke-National University of Singapore Graduate Medical School; 2Genome Institute of Singapore, Agency for Science, Technology, and Research; 3Cancer Science Institute of Singapore, National University of Singapore; 4Cellular and Molecular Research, National Cancer Centre Singapore; 5Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore; 6Department of Gastroenterology and Hepatology, National University Health System; and 7Singapore Gastric Cancer Consortium, Singapore

Gastric cancer (GC) is globally the fifth most common cancer and third leading cause of cancer death. A complex disease arising from the interaction of environmental and host-associated factors, key contributors to GC’s high mortality include its silent nature, late clinical presentation, and underlying biological and genetic heterogeneity. Achieving a detailed molecular understanding of the various genomic aberrations associated with GC will be critical to improving patient outcomes. The recent years have seen considerable progress in deciphering the genomic landscape of GC, identifying new molecular components such as ARID1A and RHOA, cellular pathways, and tissue populations associated with gastric malignancy and progression. The Cancer Genome Atlas (TCGA) project is a landmark in the molecular characterization of GC. Key challenges for the future will involve the translation of these molecular findings to clinical utility, by enabling novel strategies for early GC detection, and precision therapies for individual GC patients.

Keywords: Cancer Genetics; Cancer Genomics; Gastric Cancer.

Gastric cancer (GC) is currently the third leading cause of global cancer-related death, and particularly prevalent in Asia. Despite a steadily declining incidence, GC still causes more than 723,000 deaths a year.1 Achieving a detailed molecular understanding of GC pathogenesis is pivotal to improving patient outcomes for this complex disease. This is clearly exemplified by the TOGA Phase III trial, where GC patients with HER2/neu receptor tyrosine kinase (RTK)-positive tumors experienced clinical benefit from chemotherapy plus trastuzumab (a HER2-targeting antibody) relative to chemotherapy alone.2 In this review, we have attempted to capture the latest advances in GC molecular genetics. Special attention is paid to the recent landmark TCGA study, where almost 300 GCs were simultaneously profiled on multiple molecular platforms to identify 4 genomic subtypes: CIN (chromosomal instability), MSI (microsatellite instability), GS (genome stable), and EBV (Epstein-Barr virus).3 The comprehensive nature of the TCGA study provides an invaluable resource upon which to interpret other related GC findings.

Gastric Premalignancy and the Critical Role of Inflammation

The human stomach consists of the fundus, corpus or body, and pyloric antrum. The gastric mucosa contains three main types of glands: cardiac glands (containing mucus-producing foveolar cells), oxyntic glands (parietal cells and chief cells producing hydrochloric acid and pepsinogen respectively), and pyloric glands with mucus–secreting cells and endocrine G cells secreting gastrin. Chronic atrophic gastritis and intestinal metaplasia (IM) involving the gastric mucosa are considered important steps in GC pathogenesis (Figure 1).3 Mucosal atrophy is characterized by loss of glandular elements and replacement by metaplastic cells or fibrosis, with concomitant hypochlorhydria. IM is a pre-neoplastic lesion characterized by transformation of the gastric mucosa into an intestinal-like phenotype, replete with goblet cells and intestinal mucins. IM, associated with...
over-expression of the homeobox transcription factor CDX2, is currently classified into three subtypes: intestinal type, gastric type, and mixed gastric-intestinal type. The latter, also known as incomplete IM, is considered to confer the highest risk for GC development. Injury to the gastric mucosa has also been observed to cause metaplastic changes with spasmolytic peptide expression. This phenomenon, termed spasmolytic peptide expressing metaplasia (SPEM) or trefoil factor family 2 (TFF2) expressing metaplasia, is induced after *Helicobacter pylori* (*H pylori*) infection and chronic gastritis. Spasmolytic peptide expressing metaplasia has also been implicated as a precursor event in GC progression.

In most patients, frank GC is often preceded by several decades of chronic gastric mucosal inflammation. Chronic inflammation activates the NF-κB transcription factor, a key mediator of tumor promotion. Chronic inflammation also causes increased oxidative stress, due to reactive oxygen species and nitrosamines generated by leukocytes and macrophages which can damage proliferating cells. Moreover, the production of chemokines and cytokines may induce not only leukocyte migration but also promote carcinogenesis. Experimentally, mice with impaired immune systems such as severe combined immune deficiency (SCID) or recombinase activating gene (RAG2)-deficient mice are very susceptible to *H pylori* infection, but do not develop significant gastric disease. Such data further supports an important functional role of the host immune system in GC pathogenesis.

### Cause and Epidemiology of Gastric Cancer

Environmental factors play critical roles in GC pathogenesis, with major risk factors being *H pylori* infection, diet and smoking. High salt intake, often due to traditional diets containing salted fish, is the best documented dietary risk factor for atrophic gastritis. Recent data also suggests that iron deficiency may be a GC risk factor, as iron depletion can accelerate the progression of carcinogenesis by augmenting *H pylori* virulence. *EBV* infection is recognized as an etiological agent in 5%–10% of GCs.

*H pylori* is the most significant environmental risk factor for GC and is recognized as a class I carcinogen by the World Health Organization. Although more than 50% of the world population is infected with *H pylori*, only 1%–2% of infected people will develop GC in their lifetime. A major *H pylori* virulence factor is the cytotoxin-associated gene A (CagA), within the bacterial cag pathogenicity island (cag-PAI) which is injected into the cytoplasm of gastric epithelial cells upon microbial colonization. In transgenic mice, systemic expression of CagA has been shown to induce gastric epithelial hyperplasia, gastric polyps, gastric and intestinal adenocarcinomas.

Genetic variations in CagA, specifically present in Asian strains but not non-Asian strains, are also associated with increased chronic gastritis, gastric ulcer and gastric adenocarcinoma in human patients. In gastric epithelial cells, CagA undergoes tyrosine phosphorylation by Src kinase and activates SHP-2 (Src homology 2-containing tyrosine phosphatase). Activated SHP-2 induces the Ras-ERK pathway, a key regulator of cell growth, migration, and adhesion.

CagA also disrupts tight junctions and can target the PAR1/MARK kinase to alter apical-basolateral cell polarity, ultimately causing disorganization of the gastric mucosal architecture. CagA also has tyrosine phosphorylation-independent functions, including interactions with the Met tyrosine kinase and E-cadherin. This latter interaction disrupts binding between E-cadherin and β-catenin, leading to nuclear accumulation of β-catenin.
and subsequent activation of β-catenin-dependent transcription.22

Several clinical studies have evaluated the utility of *H pylori* eradication for GC prevention. Randomized trials in high risk populations have suggested that *H pylori* eradication is most effective in patient subgroups where premalignant lesions have not yet developed.23 For example, a *H pylori* eradication program in Matsu Island, Taiwan showed that eradication reduces the incidence of atrophic gastritis and peptic ulcer but not intestinal metaplasia.24 In a meta-analysis of 6 randomized controlled trials, *H pylori* eradication reduced the risk of developing GC by RR of 0.65 (95% CI, 0.43–0.98).25 In Japan, which has high rates of GC prevalence and *H pylori* infection, the national health insurance system approved eradication therapy for *H pylori*-related chronic gastritis in 2013.26

Smoking has been associated with a 1.5–2.5 fold increased risk of GC in both case-control and cohort studies, and this risk increases with the frequency and duration of smoking. Studies have reported that smokers have higher hazard ratios (HR) of GC in the cardia (HR, 2.86–4.10) compared to distal portions of the stomach (HR, 1.52–1.94).27 Carcinogens in tobacco smoke, notably nitrosamines and other nitroso compounds, may exert mutagenic effects, thereby increasing GC risk.28 Smoking also increases the risk for precancerous lesions such as intestinal metaplasia and dysplasia.29

*EBV* has been detected in 5%–10% of GC.13,14 A meta-analysis of 70 studies including 15,952 cases of GC demonstrated that EBV-associated GC predominantly arises in men and is found in cardia and body sites more often than in the antrum.15 A major molecular feature of EBV-associated GCs is CpG island promoter methylation of cancer-related genes.30 Expression of EBV viral latent membrane protein 2A (LMP2A) may be responsible for the promotion of DNA methylation, by inducing STAT3 phosphorylation and subsequent transcription of DNA methyltransferase 1 (DNMT1). EBV also encodes a large number of microRNAs (miRNAs),31 mostly encoded in the BHRF1 and BART regions of the genome. EBV miRNAs can repress cellular proteins, including p53 up-regulated modulator of apoptosis (PUMA), DICER1, and BIM.32 A recent study profiling 44 known EBV miRNAs in clinical samples from EBV-associated GC patients found that EBV-miR-BART4-5p plays an important role in gastric carcinogenesis through regulation of apoptosis.33

**Gastric Cancer Host Genetics**

Besides environmental agents, GC pathogenesis also involves host genetic factors. One of the first host genetic factors identified in GC involved polymorphisms in the proinflammatory gene interleukin 1-β (IL-1 β).34 A more recent study reported that combinations of single nucleotide polymorphisms (SNPs) in immune-related genes [interleukin 1-β (IL-1 β), interleukin 1 receptor antagonist (IL-1RN), tumor necrosis factor-α (TNF-α) and interleukin 10 (IL-10)] conferred a manifold increased risk of developing GC, but only in *H pylori* infected patients.35 Furthermore, a combination of high-risk host genotypes and high-risk *H pylori* genotypes greatly increased the probability of GC, up to 87-fold over baseline. These studies highlight again the critical role of host immunity in determining GC outcomes following *H pylori* infection.36 Certain high risk GC genotypes may differ in different populations, for example, a meta-analysis showed Asian high risk SNPs differ from those identified in the non-Asian population.37 Another identified host genetic factor for GC is prostate stem cell antigen (*PSCA*), which influences susceptibility to diffuse-type GC and may play a role in gastric epithelial cell proliferation.38 *PSCA* genetic variants have also been recently confirmed as influencing GC in a wide variety of populations.39

Hereditary GC syndromes are rare and account for 1-3% of GC, comprising mainly three types: hereditary diffuse gastric cancer (HDGC), gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) and familial intestinal gastric cancer (FIGC).40 HDGC is due to E-cadherin (*CDH1*) germline mutations,41 and patients with this condition are often managed with prophylactic gastrectomies.42 Besides *CDH1*, other recently identified candidate HDGC genes include *CTNNAA1*, *BRCA2* and *STK11*.43 GAPPS, first reported in 2012, is characterized by autosomal dominant transmission of fundic gland polyposis (including dysplastic lesions or adenocarcinoma or both) restricted to the proximal stomach with no evidence of colorectal/duodenal polyposis or other hereditary gastrointestinal cancer syndromes.43 Its genetic cause is as yet unknown. GC has also been associated with other hereditary cancer syndromes such as Lynch syndrome, hereditary nonpolyposis colorectal cancer syndrome, familial adenomatous polyposis, and Li-Fraumeni syndrome.40

**Gastric Cancer Pathology**

GC is morphologically, biologically, and genetically heterogeneous. Most GCs are adenocarcinomas, with lymphomas, sarcomas, and carcinoids comprising less than 5%. Adenocarcinomas are grouped by the Lauren classification into intestinal and diffuse types.44,45 In intestinal type tumors, the malignant cells form gland-like structures, unlike those in diffuse type tumors. Intestinal type GC is more common in high-risk regions while diffuse type GC is more common in low-risk areas. Diffuse type GC is more common in young patients, in whom there is a female preponderance,45 and behaves more aggressively than the intestinal type.46 Besides the Lauren classification, other histopathological classifications for GC have been proposed. For example, the WHO classification groups GC into 4 main types based on the predominant histological pattern: tubular, papillary, mucinous and signet ring cell.47

**Molecular Genetic Landscape of Gastric Cancer**

Studies from several groups over the past decade have now produced a near-comprehensive catalogue of GC-associated “driver” alterations, including gene mutations,
somatic copy number alterations (sCNAs), structural variants, epigenetic changes, and transcriptional changes involving mRNAs and noncoding RNAs (ncRNAs) (Figure 2). Certain driver alterations can also be associated with specific GC subtypes. For example, MSI is a genetic hypermutability phenomenon where microsatellite regions in tumor genomes accumulate mutations due to defective DNA mismatch repair. Next-generation sequencing (NGS) studies have shown that 15%–20% of GCs are characterized by MSI.

For ease of presentation, our review has been structured to summarize each molecular level independently.

Gene Mutations in Gastric Cancer

Like most solid epithelial cancers, GCs are often characterized by somatically-acquired mutations in various genes. Current estimates based on NGS indicate that in each individual GC (excluding hypermutated cases), there are approximately 50 to 70 nonsynonymous mutations, a mutation level comparable to colon and esophageal cancer.48 Mutated genes in GC can be broadly classified into three categories: a) high-frequency drivers displaying a high-rate of recurrence (>5%–10%) across multiple GCs; b) low-frequency drivers that are recurrently mutated in the 1%–10% range, but which still contribute to disease pathogenesis; and c) bystander/passenger mutations that arise as a consequence of underlying mutational processes such as CpG deamination, but which do not functionally contribute to tumorigenesis.

Among high-frequency drivers, TP53 is the most frequently mutated gene in GC, exhibiting aberrations in ≥50% of cases.50 Reflecting TP53’s cellular function as a guardian of genomic integrity, TP53 mutated GCs often exhibit high levels of sCNAs involving both broad chromosomal regions and focal gene regions. GCs have also been shown to exhibit mutations in other canonical oncogenes (KRAS, CTNNB1, PIK3CA) and tumor suppressor genes (SMAD4, APC).52 Reflecting the importance of RTK/RAS/MAPK signaling in GC, frequent mutations in the ERBB3 RTK and the ligand/RTK NRG1/ERBB4 genes have recently been reported.53,54 Some of these mutations appear to be enriched in specific GC subtypes, for example, EBV-positive GCs frequently exhibit PIK3CA mutations,3 while diffuse-type GCs have been observed to exhibit frequent somatic mutations in CDH1.55

Recent NGS studies of GC have also highlighted two new GC genes - ARID1A and RHOA. ARID1A, mutated in 10%–15% of GCs, encodes a component of the SWI/SNF chromatin remodeler complex. ARID1A mutations in GC are typically inactivating (eg, frameshifts, nonsense mutations), consistent with ARID1A acting as a tumor suppressor. Functional analysis suggests that ARID1A plays a role in GC cell proliferation, through the control of cell-cycle regulators CCNE1 and E2F1.57 RHOA was recently shown by multiple studies to exhibit recurrent mutations in diffuse-type/genome-stable GCs.55,56 Unlike ARID1A where mutations are dispersed throughout the gene, the RHOA4 mutations are localized to an N-terminal hot-spot region (Tyr42, Arg5 and Gly17), and are predicted to modulate downstream Rho signaling. Functional analysis of these RHOA mutations suggest that they may impart resistance to anoikis, a form of programmed cell death occurring after cellular detachment from a solid substrate.55 Clinically, the discovery of RHOA hot-spot mutations is particularly exciting, as it provides an inroad for the development of new approaches to target diffuse-type GCs, traditionally associated with extremely poor prognosis.

Besides high-frequency driver mutations, genomic studies have also uncovered a “long tail” of low frequency
driver mutations in GC, such as the gastric mucin \textit{MUC6}, \textit{BCOR} (encoding a BCL6 corepressor), \textit{FAT4} (a protocadherin), and \textit{RNF43}, a Wnt pathway regulator.\textsuperscript{3,5,5,5,7} Despite their lower mutation frequencies, these genes also likely contribute to the ability of GCs to manifest different cancer hallmarks. For example, mutations in the antigen-presenting genes \textit{HLA-B} and \textit{B2M}, observed in MSI-positive GCs, may play a role in evading host anti-tumor immune responses arising from the presentation of tumor neo-antigens.\textsuperscript{5} A key challenge for the GC field will be to develop systematic, high-throughput approaches to functionally test the effects of these mutations, either singly or in combination, on the GC cell.

NGS analyses of GCs, particularly at the whole-genome level, are also providing insights into prominent pathways and mutation signatures associated with GC development.\textsuperscript{5,5,5,6,6,6,6} For example, when all mutated genes (regardless of driver status) are considered in a collective pathway analysis, cell adhesion and chromatin remodeling are consistently highlighted as the top major disrupted pathways.\textsuperscript{5,6,6,6} Emerging evidence suggests that these processes are likely to be disrupted even in premalignant gastric tissues and early GC.\textsuperscript{5,1,6,6} A survey of mutational "signatures", referring to mutations occurring in particular sequence contexts, has highlighted specific carcinogenic and mutational processes active in GC, including microsatellite instability, CpG age-associated deamination, and the activation of cytidine deaminases such as AID and APOBEC3B.\textsuperscript{6,3,5,5,6,5} Emerging studies are now being performed to study patterns of GC tumor evolution and clonality, occurring either within the primary tumor, or between primary lesions and metastatic sites.\textsuperscript{5,5,5,6,6,6} Such results will be crucial for identifying molecular events associated with GC progression rather than initiation, and may suggest improved strategies for managing patients with advanced metastatic GC.

\section{sCNAs and Structural Variation in Gastric Cancer}

sCNAs are another major mechanism by which onco-genes and tumor suppressor genes are selectively activated or inactivated in GC. Several studies have reported genome-scale analyses to identify GC genes affected by sCNA,\textsuperscript{5,7,6,8} highlighting specific genes targeted by this mechanism. The most predominant class of genes frequently amplified in GC are those related to RTK/RAS/MAPK signaling, including \textit{HER2}, \textit{EGFR}, \textit{MET}, \textit{FGFR2}, and \textit{RAS} genes (KRAS, and to a lesser extent \textit{NRAS}).\textsuperscript{3,6,7} In total, about 30%-40% of GCs are likely to harbor a RTK/RAS/MAPK-related amplification, and therapies targeting these genes are in clinical trials, including trastuzumab (HER2), nimotuzumab (EGFR), AZD4547 (FGFR2), and onartuzumab (MET) (Table 1). Another gene frequently amplified in GC that may intersect with the canonical RTK/RAS/MAPK pathway is \textit{VEGFA}, a mediator of angiogenesis that is particularly noteworthy given the role of anti-angiogenic therapy in GC.\textsuperscript{5,6,9} In the TCGA study, EBV-positive GCs were also found to harbor frequent amplifications in \textit{PD-L1} and \textit{L2}, which are immune checkpoint inhibitor genes, and are of particular relevance as targets of immunotherapy.\textsuperscript{3}

Another class of genes targeted by gene amplification in GC are related to cell cycle control, including \textit{CCND1}, \textit{CCNE1}, and \textit{CDK6}. These cell-cycle genes are clinically relevant in two capacities. First, direct therapies are available for targets such as \textit{CDK6}, while \textit{CCND1} and \textit{CCNE1}, which encode cognate CDK partners (CDK4/6 for CCND1, CDK1/2 for CCNE1). Second, coamplification of these cell-cycle regulators with other therapeutic targets, particularly those in the RTK/RAS/MAPK pathway, may modulate responses to the therapies targeting the latter. For example, \textit{CCNE1} amplifications frequently occur together with \textit{HER2} amplifications,\textsuperscript{6,7,70} a configuration that has been linked to acquired trastuzumab resistance in breast cancer.\textsuperscript{71} \textit{HER2/CCNE1} coamplification has recently been validated as a mechanism of primary resistance against lapatinib (a dual EGFR/HER2 inhibitor) in \textit{HER2}-amplified GC.\textsuperscript{70} Knowledge of the amplification status of these cell-cycle related genes in individual GCs may thus further improve our ability to stratify patients in clinical trials employing targeted agents.

A third class of GC-amplified genes corresponds to those involved in transcriptional regulation, such as GATA factors (GATA4, GATA6), \textit{KLF5}, and \textit{MYC}. Amplifications of GATA factors and \textit{KLF5} appear to be restricted to gastrointestinal, and possibly related hepatobiliary malignancies.\textsuperscript{72,73} These transcription factors (TFs) are expressed in the developing and normal stomach,\textsuperscript{74} and may act as “lineage-survival” factors functioning to reawaken early developmental programs to drive GC tumorigenesis. Interestingly, recent chromatin immunoprecipitation (ChIP) sequencing studies provide evidence that these amplified TFs, may in fact collaborate with one another to regulate common downstream promoters in genes associated with GC development.\textsuperscript{73} Another transcription factor amplified in GC is \textit{OCT1}, a regulator of ERK signaling.\textsuperscript{75} Although traditionally considered to be “undruggable” by the pharmaceutical industry, novel chemical biology technologies for disrupting such oncogenic TFs are in development.\textsuperscript{76}

Besides amplifications, genomic deletions also frequently occur in GC, involving tumor suppressor genes such as \textit{WWOX}, \textit{RB1}, \textit{PARK2}, \textit{FHIT}, and \textit{CDKN2A/B}.\textsuperscript{6,7} Somatic deletions in \textit{CDH1} have been associated with poor patient prognosis,\textsuperscript{77} and genomic deletions involving nonprotein coding genes such as \textit{mir-101a} have been reported, which can cause upregulation of the oncogenic histone methyltransferase \textit{EZH2}.\textsuperscript{78} Interestingly, while most genomic deletions are often interpreted as loss-of-function events, a recent study has shown that the phosphodies-terase \textit{PDE4D}, frequently targeted by microdeletions in several cancers including GC, is associated with a constitutively active PDE4D isoform that promotes tumor growth.\textsuperscript{79} It is thus possible that, as sequencing technologies improve, detailed fine mapping of genomic deletion break-points will highlight similar unexpected examples, prompting the revisiting of these “deleted” genes as tumor suppressors.

Genomic rearrangements, either balanced or associated with copy number changes, can also result in fusion genes,
creating new chimeric protein products (eg, BCR-ABL in chronic myelogenous leukemia), or causing high expression of one gene through hijacking of the partner’s promoter (eg, TPMPRSS2-ERG in prostate cancer). An RNA-sequencing study in 2010 revealed for the first time the presence of rare RAF-fusion genes in GC and prostate cancers (eg, AGTRAP-BRAF), and preclinical models expressing these RAF fusions were shown sensitive to treatment by sorafenib, originally designed as a RAF inhibitor. The second fusion gene identified in GC was CD44-SLC1A2, which juxtaposes the SLC1A2 glutamate transporter against the CD44 promoter. Tumors expressing this gene fusion expressed high levels of SLC1A2 which may facilitate the acquisition of glutamate by the cancer cell to fuel cancer metabolism. Subsequently, CD44-SLC1A2 and related AP1P-SLC1A2 fusions have been observed in colon cancer, indicating that such fusions may be restricted to gastrointestinal-type cancers. ROS1 rearrangements in GC have been recently reported, and in the TCGA study, CLDN18-ARHGAP26 fusions were observed in genome-stable/diffuse type GC. Interestingly, ARHGAP26 is a member of the Rho-effector pathway, and the CLN18-ARHGAP26 fusions are mutually exclusive to RhoA mutations and CDH1 mutations, all pointing to a common dependence on cell-adhesion related Rho signaling in diffuse-type GC. Indeed, recent functional data has confirmed that CLDN18-ARHGAP26 expression in epithelial cells can reduce cell-cell and cell-extracellular matrix-adhesion, while contributing to epithelial-mesenchymal transition (EMT). Taken collectively, these examples attest to the presence of gene fusions in GC, which due to their cancer-specific nature, may be exploited for improved diagnostics or therapeutics.

**Epigenetic Alterations involving DNA Methylation and Chromatin**

Epigenetic aberrations are also emerging as important players in GC pathogenesis. These can occur at the level of DNA methylation, where methyl groups are attached to cytosine bases at CG motifs; or histone modifications where specific residues in histones are methylated, acetylated, or phosphorylated. There is great interest in studying how environmental risk factors for GC may modulate the gastric epigenome. For example, H pylori has been shown to infect and induce widespread DNA methylation changes in stomach epithelial cells. Such “epigenetic field defects” are likely to contribute to an individual’s risk of developing GC. The epigenetic field model has recently been supported by a recent prospective clinical trial.

There is a rich body of GC literature relating promoter DNA methylation to the transcriptional silencing of tumor suppressor genes (CDH1, RUNX3, p16, and hMLH1). Alterations in DNA methylation can also influence GC progression and treatment response. For example, methylation-associated silencing of PLA2G2A, a secreted phospholipase, may play a role in driving aggressive GC, and methylation of SULF2 and BMP4 have been linked to chemotherapeutic responses. Beyond individual genes, genome-wide studies have also revealed large scale patterns of DNA methylation in GC. Multiple studies have confirmed that certain GCs can exhibit a CpG island methylation phenotype (CIMP), characterized by genome-wide methylation of CpG islands. Patients with CIMP type GCs tend to be younger, with less-differentiated tumors. Pathway analysis reveals that genes targeted by CIMP in GC are enriched in genes corresponding to PRC2-targets in embryonic stem cells, suggesting a reawakening in CIMP tumors of a nascent stem cell program. Interestingly, in preclinical assays, CIMP-positive GCs appear to exhibit heightened sensitivity to DNA methyltransferase inhibitors such as 5’aza.

As described by earlier studies and confirmed by TCGA, EBV-positive GCs also exhibit exceptionally high DNA methylation levels. Indeed, among all tumor types studied by TCGA to date, EBV-positive GCs appear to exhibit the highest DNA methylation levels, surpassing even CIMP tumors. It is possible that such hypermethylation may represent a cellular reaction to viral infection. Emphasizing the importance of tumor immunity in this GC subtype, EBV-positive GCs also typically exhibit a high lymphocytic

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Prevalence in GCa</th>
<th>Representative Therapies</th>
<th>Clinical Statusb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERRB2</td>
<td>Amplification/overexpression</td>
<td>10%–20%</td>
<td>Traztuzumab/trastuzumab emtansine</td>
<td>Approved/Phase III</td>
</tr>
<tr>
<td>VEGF</td>
<td>Overexpression</td>
<td>~50%</td>
<td>Ramucirumab</td>
<td>Approved</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>Overexpression</td>
<td>40%–50%</td>
<td>Bevacizumab/regorafenib</td>
<td>Phase III (negative)/phase II</td>
</tr>
<tr>
<td>EGFR</td>
<td>Amplification/overexpression</td>
<td>6%–27%</td>
<td>Cetuximab/nimotuzumab</td>
<td>Phase II/III</td>
</tr>
<tr>
<td>MET</td>
<td>Amplification/overexpression</td>
<td>5%–40%</td>
<td>Onartuzumab/AMG337</td>
<td>Phase II/III</td>
</tr>
<tr>
<td>FGF2</td>
<td>Amplification/overexpression</td>
<td>4%–12%</td>
<td>AZD4547/dovitinib</td>
<td>Phase II</td>
</tr>
<tr>
<td>ATM</td>
<td>Loss (Protein)</td>
<td>60%</td>
<td>Olaparib</td>
<td>Phase III</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>Mutation</td>
<td>5%–10%</td>
<td>Everolimus/GDC0068</td>
<td>Pre-clinical/early trials</td>
</tr>
<tr>
<td>CDK4/6</td>
<td>Amplification</td>
<td>6%–15%</td>
<td>LEE001</td>
<td>Phase II</td>
</tr>
<tr>
<td>PD-L1/P2</td>
<td>Amplification/overexpression</td>
<td>15% of EBV-positive GC</td>
<td>MPDL3280A</td>
<td>Preclinical/phase I</td>
</tr>
<tr>
<td>MSI</td>
<td>Mutation</td>
<td>15%–20%</td>
<td>Pembrolizumab (anti-PD1)</td>
<td>Phase II</td>
</tr>
<tr>
<td>ARID1A</td>
<td>Mutation</td>
<td>8%–10%</td>
<td>EZH2 inhibition</td>
<td>Preclinical</td>
</tr>
</tbody>
</table>

**GC, gastric cancer; EBV, Epstein-Barr virus; MSI, microsatellite instability.**

*Most ongoing trials do not involve biomarker-selected GC patients.*
infiltrate, and (as described above) also frequently exhibit PD-L1 and PD-L2 amplification. Another recent study has recently reported additional mutated genes in EBV-positive GC, which may contribute to tumor neo-antigen formation in this unique and intriguing cancer subtype. In contrast to DNA methylation, our understanding of histone modifications in GC is still embryonic. An earlier study assessing patterns of the H3K27me3 repressive mark found over a hundred genes exhibiting H3K27me3 differences between GCs and matched normal tissues. More recently, a study employing a nano-scale chromatin immunoprecipitation techniques to profile multiple histone marks in GCs and matched normal tissues, found evidence of widespread alternative promoter usage and deregulated enhancer elements, including an alternative isoform of the MET RTK that produces a MET protein variant lacking the SEMA N-terminal regulatory domain. Future work will be focused on understanding the mechanisms underlying such regulated regulatory elements, and how they might be harnessed for predictive or prognostic potential.

**Gastric Cancer Transcriptome—Gene Signatures and Alternative Splicing**

Altered regulation of gene expression programs are critical to the ability of tumors to express different cancer hallmarks. In early 2000, early microarray studies already revealed a staggering diversity of expression programs operative in GC, associated with different histological subtypes, tumor status, and clinical traits. In the past decade, groups have deployed more mature transcriptomic technologies, on larger patient cohorts, to identify robust gene expression "signatures" that can subtype GC in a clinically relevant fashion. Such GC transcriptome studies have typically fallen into two categories: a) unsupervised, where GC subtypes are discovered based on molecular data in an unbiased fashion and then correlated to clinical data; or b) supervised, where clinical data is directly correlated to molecular data to identify highly associated expression patterns. Among unsupervised studies, one report, starting from a panel of GC cell lines, identified two "intrinsic" subtypes of GC (G-INT/G-DIF), that when mapped to primary GCs correlated with Lauren’s classification, and was both prognostic and predictive to 5-FU related chemotherapies. This classification has recently been extended to three subtypes - proliferative, mesenchymal, and metabolic, where cell lines with mesenchymal features were found to be more responsive to PI3KCA/mTOR inhibitor compounds. Other unsupervised classifications have been proposed, including a recent 4-subtype system by the Asian Cancer Research Group (ACRG), a collaboration between academia and industry. In complement to these, groups performing supervised analyses have also reported specific gene signatures for directly predicting GC patient prognosis. One study reported a 6-gene signature that could be used to compute an individualized "risk-score" for predicting GC prognosis. In another study from Korea, investigators explored expression patterns of >1000 GCs to identify an 8-gene signature capable of predicting the prognosis of Stage II GC patients. This is a highly clinically-relevant question, as Stage II patients identified as “poor prognosis” by the gene signature might be prescribed adjuvant chemotherapy, while "good prognosis" patients might be treated with surgery alone.

Transcriptomic information can also be exploited as a powerful discovery method to identify deregulated pathways in GC. One study exploring this approach reported that GCs exhibiting combinations of activated oncogenic pathways were associated with poor prognosis relative to GCs with activation of single pathways. Investigators have also used RNA-sequencing data of GCs from different countries to identify the AMPK/HNF4α/Wnt5α pathway as a targetable oncogenic pathway in early stage GC. However, it should be noted that while transcriptomes are highly dynamic, almost all GC transcriptomic studies to date have relied upon analyzing GCs obtained at a single time-point, usually upon surgical resection. There is thus great interest in understanding gene expression (and other molecular) changes in GCs during temporal progression or treatment. Demonstrating the power of this temporal approach, one highly notable study used paired endoscopic biopsies to study gene expression changes in GC patients prior to and after developing resistance to 5-FU/platinum therapies. They identified an acquired-resistance signature comprising genes related to cell survival (mTOR pathway), DNA repair, and embryonic stem cell biology. Data from such efforts, while still scant in the literature, will prove vital in attempts to overcome chemotherapy resistance in GC.

Transcriptomic analysis is also revealing new molecular features of GC biology, above and beyond those provided through the study of DNA. One especially exciting field of exploration is alternative splicing, where different transcript isoforms of the same gene may play distinct functional or oncogenic roles. One of most famous examples of alternative splicing in GC involves the CD44 gene, which encodes a cell-surface glycoprotein that binds hyaluronic acid and mediates a diversity of cell-cell interactions. Recent data has revealed that GC stem cells express a specific CD44 transcript variant (CD44v8-10) that can be used to enrich for tumor-initiating cells. Functional analysis of this variant suggests that its role is to act as a copartner of the xCT cysteine transporter, to regulate oxidative stress. Another recent example of alternative splicing in GC, revealed through integrated exome/transcriptome analysis, involves the selective use of the ZAK kinase isoform TV1 in tumors, which is required to maintain cell proliferation in experimental systems. Alternative splicing events can also cause production of fusion genes in the absence of genome rearrangements, through the process of “read-through” transcription. One example of this process is PPP1R1B-STAR, involving two genes adjacent to one another in the human genome, which fuses exon 6 of PPP1R1B to exon 2 of STAR. Interestingly, functional analysis suggests that gene products arising from such “read-through” fusions may be enriched in gastric tumors related to matched normal controls, and may contribute towards certain pro-oncogenic features.
Non-coding RNAs, miRNAs and Beyond

Non-coding RNAs represent yet another active player in GC pathogenesis, involving different RNA classes such as miRNAs and long intergenic noncoding RNAs (lincRNAs). miRNAs are small RNAs (~22 nucleotides) that can bind to target gene transcripts to induce the latter's transcript or protein downregulation. Studies from multiple groups have now produced a growing list of >200 miRNAs with potential roles in GC development, progression, and treatment response. Examples of oncogenic miRNAs include miR-21, miR-27a, and miR-130b; which are overexpressed in GC tumors and cell lines relative to normal controls. Oncogenic miRNAs typically function to target and downregulate tumor suppressor genes, for example, miR-21 has been reported to target the tumor suppressor genes PTEN and PDCD4, while miR-130b has been shown to regulate RUNX3. Reciprocally, tumor suppressor miRNAs such as miR-375, miR-486, miR-29c, and miR-101 are downregulated in GC, typically by DNA methylation or genomic loss, and are required to silence genes with oncogenic potential such as PDK1, OLFM4, ITGB1, and EZH2 respectively. Some miRNAs may also contribute to GC pathogenesis by functioning as key components of oncogenic signaling pathways, as shown by recent data implicating mir-365 in PTEN/Akt signaling, or by responding to external carcinogens (H pylori) to stimulate cellular proliferation (eg, mir-210). Work is also ongoing to characterize the role of miRNAs in tumor progression and drug resistance, with numerous studies reporting miRNA expression patterns associated with patient outcome. For example, miRNAs in the mir-200 family play a key role in regulating EMT via EMT regulators such as ZEB1, and mir-200 downregulation is associated with poor prognosis.

In contrast to miRNAs, our current knowledge of the role of lincRNAs in GC is still immature, however recent studies examining expression of lincRNAs such as HOTAIR and GAPLINC in GC suggest an important role for this RNA class in GC development and progression. Identifying the most prevalent deregulated lincRNAs in GC, and deciphering their functional and mechanistic roles in regulating gene expression and cancer behavior, will clearly be highly fertile area for future GC research.

Beyond the Cancer Cell: Contributions From the Tumor Microenvironment

Gastric tumors are composed not just of cancer cells but also other cell populations including stromal cells, immune cells, and blood vessels. Far from being innocent bystanders, studies are now demonstrating that these “noncancer” compartments are also likely to play important roles in GC disease progression and aggressiveness. Several studies have now confirmed that GCs with a high stromal content, determined either by gene expression analysis or histopathological evaluation, are associated with poor prognosis. Similar findings have also been reported in other tumor types, such as breast, lung and esophageal cancers. The mechanisms underlying the role of the tumor stroma in supporting GC growth are likely multifactorial, ranging from providing a suitable extracellular matrix to maintain cancer cell growth, active cell-cell interactions via cancer-associated fibroblasts which can secrete growth-stimulatory molecules such as IL-6, and providing a physical barrier inhibiting chemotherapy from reaching target malignant cells. Major pathways implicated in the tumor stroma include TGFβ signaling and EMT-related signaling events, pointing to complex interactions between cancer cells and their nonmalignant neighboring cellular populations.

Studies are now emerging at attempting to target the tumor stroma for therapy.

Another intra-tumor cellular population vitally important to GC development are cells regulating tumor vasculature and immunity. GCs have been shown to secrete pro-angiogenic compounds such as VEGF, CXCL1, and Ang-2 to stimulate blood vessel formation. In GC, multiple clinical trials targeting various aspects of the tumor angiogenesis axis have been conducted. In the AVAGAST trial, the combination of bevacizumab (an antibody targeting VEGFA) and chemotherapy did not extend overall survival compared to patients treated with chemotherapy alone. Recently however, the RAINBOW and REGARD trials evaluating ramucirumab, a VEGFR2-targeting antibody, reported an improvement in overall survival in patients with advanced GC. Interestingly, for at least two of these studies (AVAGAST and RAINBOW), subsequent subset and biomarker analysis revealed that GC patients from non-Asian countries tended to receive clinical benefit from the addition of bevacizumab, compared to patients from Asian countries. It is thus possible that GC populations in different countries (specifically Asian and non-Asian countries) may respond differently to anti-angiogenic and potentially other therapies.

Finally, immune cells of many different types are also frequently observed in primary GCs, and certain subtypes of GC (eg, EBV-positive GC and MSI-positive GC) are well-known to be associated with a high lymphocytic infiltrate. Depending on their specific identities and immune functions, such tumor-infiltrating immune cells may play pro- or anti-tumorigenic roles. For example, GCs with high levels of regulatory T cells (FOXP3-positive) have been reported to exhibit good prognosis, while those with high levels of IL-17 producing CD8-positive T cells (cytotoxic T cells) are associated with poor prognosis. The role of tumor immunity in cancer progression is extremely broad, and readers are directed to other excellent reviews dedicated to this subject. Interestingly, two recent studies have also reported that Asian and non-Asian GC populations may differ in their regulation of tumor immunity factors at the genetic and gene expression level, which may impact region-specific effects on therapy outcome and prognosis. Given the intimate relationship between immune cells and tumor vasculature, it will be interesting to test if such differences might
Conclusions and Challenges

Our knowledge of GC molecular pathogenesis has considerably evolved over the years, but much remains unknown. In terms of basic understanding, the role of novel molecular entities such as lincRNAs and chromatin alterations in GC remains to be unraveled. DNA sequencing studies have also revealed a complex community of non-cultivable microbiota in the human stomach beyond H pylori and EBV, and it will be critical to understand how the gastric “microbiome” may interact with both H pylori and the host immune response to influence gastric carcinogenesis. In the early detection field, molecular findings may facilitate new approaches for diagnosing GC early, by identifying high-risk individuals through the molecular characterization of preneoplastic lesions or even blood-based biomarker assays. Finally, therapeutic strategies remain to be developed to target specific GC subtypes (particularly diffuse type GC), based upon their somatic driver alterations or tumor-associated cell compartments (eg, stromal cells and tumor-infiltrating immune cells). The next five years will undoubtedly see a further explosion of findings in the GC field, which will be essential to combat this deadly disease.

Supplementary Material
NOTE: The first 50 references associated with this article are available below in print. The remaining references accompanying this article are available online only with the electronic version of this article. To access the supplementary material accompanying this article, visit the online version of Gastroenterology at www.gastrojournal.org, and at http://dx.doi.org/10.1053/j.gastro.2015.05.059.

References

22. Murata-Kamiya N, Kurashima Y, Teishikata Y, et al. Helicobacter pylori CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes in-


Received March 25, 2015. Accepted May 20, 2015.

Reprint requests
Address requests for reprints to: Patrick Tan, Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, 8 College Road, Singapore 169857. e-mail: gmstanp@duke-nus.edu.sg; fax: (+65) 6226 5294; or Khay Guan Yeoh, Department of Medicine, Yong Loo Lin School of Medicine, 1E Kent Ridge Road, NUHS Tower Block, Level 11, Singapore 119228. e-mail: khy_guan_yeoh@nuhs.edu.sg; fax: (+65) 6778 5743.

Acknowledgements
We thank Beatrice Tan, Jennie Wong, Tze Wei Chua, and Feng Zhu for assistance with manuscript preparation and Wei Peng Yong and Matthew Ng for valuable feedback. We thank Prof Teh Ming and Dr Supriya Srivastava, Dept of Pathology, National University of Singapore for supplying the histopathology images for Figure 1. Due to length restrictions, we apologize to those clinicians and scientists in the GC community whose work we may have inadvertently missed.

Conflicts of interest
The authors disclose no conflicts.

Funding
This work was supported by the Singapore Gastric Cancer Consortium, funded through the National Medical Research Council/National Research Foundation Translational and Clinical Research Flagship Program (NMRC/TCR/009-NUHS/2013).
Supplementary References


