

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

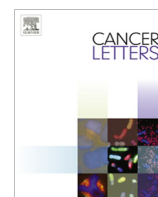
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Cancer Letters

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

Mini-review

Identifying molecular drivers of gastric cancer through next-generation sequencing

Han Liang^{a,*}, Yon Hui Kim^{b,*}^a Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030, USA^b New Experimental Therapeutics Branch, National Cancer Center, Korea, 323 Ilsan-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do 410-769, Republic of Korea

ARTICLE INFO

Keywords:

Somatic mutation
Gastric cancer
Gene fusion
Therapeutic target
Next-generation sequencing
RNA sequencing

ABSTRACT

Gastric cancer is the second most common cause of cancer-related death in the world, representing a major global health issue. The high mortality rate is largely due to the lack of effective medical treatment for advanced stages of this disease. Recently next-generation sequencing (NGS) technology has become a revolutionary tool for cancer research, and several NGS studies in gastric cancer have been published. Here we review the insights gained from these studies regarding how use NGS to elucidate the molecular basis of gastric cancer and identify potential therapeutic targets. We also discuss the challenges and future directions of such efforts.

Published by Elsevier Ireland Ltd.

1. Background

Gastric cancer refers to a tumor that originates in any part of the stomach. Currently the fourth most common malignancy in the world, gastric cancer is second to lung cancer as the leading cause of cancer mortality, causing about 800,000 deaths per year [1,2]. The incident rate of gastric cancer appears to vary greatly across geography, ethnicity and gender [3]. The mortality rate attributed to gastric cancer has recently declined in North America and Europe; however, gastric cancer remains the most common cancer type in countries of East Asia [4], representing the highest cancer incident rate among males and females in Korea, the most common cancer among females in China, and the most common cancer among males in Japan [5].

Conventional treatment options for gastric cancer include surgery, chemotherapy, radiation therapy, or combinations thereof known as multimodality therapy. If gastric cancer is diagnosed in an early stage, curative treatment can be achieved through complete surgical removal of the tumor tissue [6]. However, because gastric cancer causes few symptoms in its early development, a diagnosis is usually made after the cancer reaches an advanced stage. Moreover, even after having a gastric tumor surgically removed, many patients will experience disease recurrence and die within a few months to years. The 5-year relative survival rate of gastric cancer has not improved during the past 35 years, and remains stubbornly at 20–30%. Thus, the high mortality rate attrib-

uted to gastric cancer is due to the lack of both early detection methods and effective medical treatment for advanced stages of the disease [6].

Gastric cancer is characterized by a high level of biological heterogeneity, with each patient exhibiting a distinct genetic and molecular profile [7]. Histologically, the majority of gastric malignancies are adenocarcinomas that can be further categorized as diffuse (poorly differentiated) or intestinal (well-differentiated) types, each with distinct epidemiological and genetic patterns, and the former subtype carrying a worse prognosis [7–9]. Etiologically, gastric cancer is associated with the combined effects of environmental factors and susceptible genetic variants, including the accumulation of genetic and epigenetic alterations [7,8]. Diet and lifestyle factors have been shown to significantly affect gastric cancer risk, with greater risk associated with tobacco smoking and obesity [10]. The infectious agent *Helicobacter pylori* is closely related to the most common types of gastric adenocarcinoma [11], and is the main risk factor in 65–80% of gastric cancer cases. *H. pylori* induces generalized mutations and genomic instability in the host DNA [12], and this possibility may increase the diversity of oncogenic mechanisms in gastric cancer.

Given the diverse routes to gastric oncogenesis, it is unlikely that a “magic bullet” treatment exists. Therefore, targeted therapy based on the biology of the individual patient is highly attractive. This involves first identifying the biological molecules that contribute to the development or maintenance of the tumor, then specifically targeting those oncogenic mechanisms with an anti-tumor treatment regimen. Such therapies attempt to inactivate specific oncogenic mechanisms that are critical to the survival of tumor cells, while sparing normal gastric cells, thereby maximizing the benefits and minimizing the side effects. In early 2010, trast-

* Corresponding authors. Tel.: +1 713 745 9815; fax: +1 713 563 4242 (H. Liang), tel.: +82 31 920 2519; fax: +82 31 920 2542 (Y.H. Kim).

E-mail addresses: hliang1@mdanderson.org (H. Liang), yhkim@ncc.re.kr (Y.H. Kim).

uzumab, a humanized monoclonal antibody directed against the extracellular domain of the HER2/neu receptor, was approved for the treatment of gastric cancer. Trastuzumab became the first targeted agent to be used in combination with chemotherapy as first-line treatment of ERBB2-positive advanced and metastatic gastric cancer [13,14]. A relatively large proportion of gastric tumors (2–27%) harbor ERBB2 amplifications and may respond to this treatment regimen [15,16]. Several molecular targeted agents associated with a survival advantage in other cancer types are now under clinical investigation for the treatment of gastric cancer, including inhibitors of EGFR, MET, FGFR, VEGF and PI3K [17,18]. For example, in a single-arm phase II clinical trial for patients with gastric and gastroesophageal adenocarcinoma, Shah et al. demonstrated that the addition of bevacizumab, a monoclonal antibody against VEGF, to a therapeutic regimen of irinotecan and cisplatin achieved a median time-to-disease progression of 8.3 months, which was an improvement of 75% over that of the historical control regimen [19]. Pinto and colleagues demonstrated a response rate of 44.1% and median time-to-disease progression of 8 months in advanced gastric cancer using an irinotecan-based regimen in combination with EGFR-targeting cetuximab [20]. However, these targeted therapies, whether approved or in clinical development, have focused on a few well-known oncogenic genes, and their efficacy in gastric cancer is not yet known. Developing novel therapeutic strategies to specifically target gastric cancer requires a comprehensive dissection of the molecular mechanisms underlying the initiation and progression of this disease.

2. Brief overview of next-generation sequencing technology

Next-generation sequencing (NGS) is a powerful technology for elucidating the pathogenesis of human cancer and identifying potential therapeutic targets [21]. NGS is also known as second-generation sequencing, which describes its relation to capillary-based Sanger sequencing, the first-generation sequencing technology [22]. NGS includes the various implementations of cyclic-array sequencing [23]. Among the most widely used commercial NGS platforms are the 454 Genome Sequencer (Roche Applied Science), Solexa technology (Illumina) and the SOLiD platform (Life Sciences). Although some technical differences exist, the concept of these NGS platforms is quite similar: the sample DNA is first randomly fragmented, followed by *in vitro* ligation of common adaptor sequences, thereby forming a library; then an array of millions of spatially fixed PCR colonies are generated, each consisting of many copies of a single library fragment; and finally, enzyme-driven biochemistry and imaging-based data processing is performed in parallel. Thus, given a sample, NGS generates millions of short reads in an extremely efficient fashion. In the downstream analysis, the short read sequences are mapped to the source genome to generate a nucleotide-resolution read distribution, from which various biological features about the molecules in the sample can be inferred. Depending on the type of input materials, the major NGS applications include DNA sequencing (DNA-seq), RNA sequencing (RNA-seq) and chromatin immunoprecipitation sequencing (ChIP-seq).

The initial applications of NGS focused on sequencing the genomes of individual people or transcriptomes of various tissues. The more recent applications of NGS to study cancer biology [24] have revolutionized cancer research in several ways. First, NGS is able to generate unbiased, comprehensive rather than biased, limited catalogs of various aberrations in the cancer genomes. For example, pre-NGS sequencing studies on cancer somatic mutations examined only a small set of well-known cancer genes, such as TP53, EGFR and KRAS; and these focused studies had little power to discover novel cancer genes [24]. Second, unlike “analog” signals in earlier sequencing technology, the NGS signaling is based on the

count of the short reads. This digital-like signal is extremely good for quantification [25]. Thus, high sequence coverage provides sufficient information to detect biological signals in heterogeneous samples such as cancer tissues. Finally, with the commercialization of NGS platforms, the cost of DNA sequencing has decreased dramatically, reaching a level of ~\$1,000 per personal genome [26]. This low cost has facilitated the clinical applications by incorporating personal genomic information.

A handful of NGS studies in gastric cancer have been published over the last two years (summarized in Table 1). In this article, we review the insights gained from these studies regarding the use of NGS to identify potential therapeutic targets in gastric cancer, and discuss the challenges and future directions of such efforts.

3. Targeted DNA-seq in gastric cancer

Although the first applications of NGS in cancer were almost all whole-genome sequencing studies [24], targeted DNA-seq has become a more popular approach for identifying somatic mutations in cancer genomes. Through a DNA capture technique based on hybrid selection, targeted DNA-seq focuses on a proportion of the genome, such as exons. Although this approach misses some aberrations that could be detected in whole-genome sequencing, such as copy number alterations, interchromosomal rearrangements and inversions, targeted DNA-seq represents a cost- and resource-efficient approach, allowing for the identification of somatic mutations of greatest potential interest. Given the mutations detected in a tumor sample, the key goal is to distinguish “driver” mutated genes from “passenger” mutated genes. The somatic mutations of driver genes are those that play critical roles in cancer pathophysiology and thus represent high-quality targets; whereas those from passenger genes result from the underlying genomic instability associated with cancer and have little phenotypic effects. In general, driver genes are expected to harbor more somatic mutations than would occur randomly. Since the mutation rate in the cancer genome depends on various factors, such as coverage, base composition, gene expression [27], and replication time [28], it is essential to take these factors into consideration when assessing the statistical significance of mutations observed in a gene relative to the “background rate.” Importantly, it would be desirable to experimentally validate the functional impact of mutated genes using functional assays such as cell viability or cell transformation assays [29].

Wang and colleagues published the first exome-sequencing study in gastric cancer [30], sequencing 22 matched pairs of gastric cancer and healthy gastric tissue. Using a driver gene score, they identified 20 genes as top candidate drivers, including previously known drivers such as TP53, PTEN and CTNNA1. One major finding from the study is the high mutation frequency of ARID1A, a key member of the SWI-SNF complex that has been associated with frequent mutations in several other cancer types. Interestingly, Wang et al. also found chromatin modification and cell junction organization pathways to be the most perturbed pathways and 59% of the gastric cancer samples under survey to have mutations in histone-modifying proteins. Then, using Sanger sequencing to further evaluate ARID1A in additional samples of gastric cancer, they found that the somatic mutation rate and spectrum of ARID1A significantly varies among different molecular subtypes of gastric cancer; and clinically, alterations in ARID1A are associated with better prognosis in a stage-independent manner. Interestingly, ARID1A was also identified as a driver gene in a later exome-sequencing study in which Zang et al. sequenced 15 matched sample pairs of gastric cancer and performed similar computational analyses [31]. Importantly, they provided direct functional evidence supporting a tumor suppressor role of ARID1A: knockdown mediated

Table 1
Next-generation sequencing studies in gastric cancer.

Study	Method	Samples	Aberration type	Main findings
Wang et al. [30]	Whole-exome sequencing	22 tumor and matched normal pairs	Point mutations, small indels	Frequent inactivating mutations in ARID1A
Zang et al. [31]	Whole-exome sequencing	15 tumor and matched normal pairs	Point mutations, small indels	Frequent inactivating mutations ARID1A, FAT4
Zang et al. [32]	Whole-kinome sequencing	14 GC cell lines 3 tumor and normal pairs	Single nucleotide variations, gene fusions, and copy number variations	Recurrent inactivating mutations in MAP2K4, gene fusions involving CDK12 and ERBB2
Holbrook et al. [33]	Targeted DNA sequencing (384 genes)	44 tumor samples (36 with matched normal)	Point mutations	Genetic alternation in the WNT, Hedgehog, cell cycle, DNA damage and epithelial-to-mesenchymal transition pathways
Kim et al. [40]	Whole transcriptome RNA-sequencing	24 tumor samples and 6 normal samples	mRNA expression, miRNA expression, recurrent somatic mutations	AMPK α 2 as a potential therapeutic target in Asian patients
Riberiro-dos-Santos [42]	Small RNA Sequencing	Healthy gastric tissue	-	15 most highly expressed miRNAs in gastric tissue
Li et al. [43]	Small RNA sequencing	One pair of tumor and normal samples	microRNA expression	Biased selection of arm miRNAs from the same pre-miRNAs

by ARID1A shRNA or siRNA increased the cell proliferation in the cell lines with wild-type ARID1A. In contrast, in the ARID1A-deleted lines, no such effects of shRNA-mediated silencing were detected, and re-expression of ARID1A suppressed the cell proliferation [31]. These studies thus highlight the importance of chromatin remodeling genes in gastric tumorigenesis. In addition to ARID1A, Zang et al. identified FAT4, a member of the E-cadherin family, as a strong candidate driver gene. Using the same experimental approach in cell lines as done previously for ARID1A, they showed that FAT4 functions as a tumor suppressor [31]. Fig. 1 shows a comparison of top candidate driver genes identified by these two exome-sequencing studies.

Although the exome sequencing studies provided an unbiased way to assess the mutation patterns in all protein-coding genes (~20,000) in the human genome, novel candidate driver genes thereby obtained are often not good candidates for drug discovery. This is because (i) the driver gene itself is not readily targeted by a drug or (ii) no sufficient prior knowledge has been accumulated about the genes, making further functional or clinical investigation very difficult. Therefore, using DNA-seq to target a subset of “drug-gable” or “important” genes becomes a more effective way to discover therapeutic targets. Two such studies have been recently published. Zang et al. characterized the protein-coding regions of

537 kinases in 14 commonly studied cell lines using NGS, and detected more than 300 novel kinase SNVs [32]. A family-wise analysis further revealed a significant SNV enrichment in MAPK-related genes. In particular, siRNA knockdown and overexpression experiments identified MAP2K4 as a tumor suppressor gene in this disease. Holbrook et al. performed targeted DNA sequencing on 384 genes belonging to various pathways known to be important in cancer [33]. Their study revealed mutations in the RAS/RAF/MEK/ERK pathway, PI3K/AKT pathway, WNT pathway, and the hedgehog pathway, suggesting novel therapeutic opportunities. For example, KRAS G12D mutations imply a treatment option for MEK inhibitors [34]; and inhibitors of the hedgehog pathway, such as GDC-0449, are in clinical development as therapeutic targets [35]. In addition, the mutations identified in the study also suggested novel therapeutic targets such as the thyrotropin receptor (TSHR) and the Rho-associated coiled-coil containing protein kinases (ROCK1 and ROCK2).

4. Transcriptomic RNA-seq in gastric cancer

Before the NGS era, hybridization-based microarrays were the main approach for transcriptomic studies in gastric cancer. Using miRNA expression microarrays, Volinia et al. [36] and Ueda et al. [37] evaluated aberrant miRNA expression signatures in gastric tumor samples from Italian and Japanese patients, respectively; and Cui et al. used exon arrays to identify differentially expressed mRNAs as potential biomarkers in Chinese patients [38]. In recent years, RNA-seq has become a revolutionary approach for transcriptome profiling, supplanting microarrays with rapid speed [25]. Compared with microarrays, NGS-based RNA-seq has several advantages in cancer transcriptome profiling. First, RNA-seq provides an unbiased approach to profiling all the transcribed molecules in a sample, which is not limited to the previously known or annotated transcripts. Second, RNA-seq accommodates a large dynamic range of expression levels, thereby allowing for a much more accurate quantification of genes with very low or high expression levels [25]. Third, in addition to gene-level expression quantification, RNA-seq can detect other types of transcriptional signals, including alternative splicing, transcriptional starts/stops, gene fusion, and expressed alleles [39].

We recently published the first comprehensive RNA-seq study in gastric cancer. We applied a whole-transcriptome sequencing approach to 24 samples of gastric tumors and six noncancerous tissue specimens obtained from Asian patients, generating 680 million informative short reads to quantitatively characterize gene expression [40]. We used two sequencing protocols to respectively

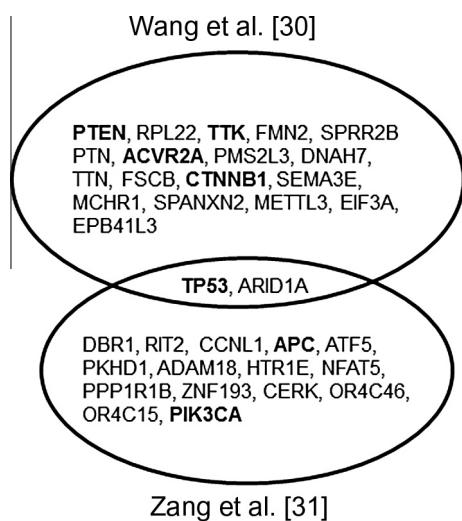


Fig. 1. A comparison of most significantly mutated genes identified in the two whole-exome sequencing studies. Top significantly mutated genes ($FDR < 0.2$) were obtained from each study, respectively; and genes previously known to associate with gastric cancer are shown in bold.

sequence the transcripts of different lengths. Beyond the conventional analysis, one unique aspect of the study is that we developed a multilayered integrative analysis to identify various types of transcriptional aberrations associated with different stages of gastric cancer, including differentially expressed mRNAs, recurrent somatic mutations, and key differentially expressed microRNAs. Through this computational approach, we identified the central metabolic regulator AMPK α 2 (PRKAA2) as a potential functional target in Asian gastric cancer. This gene shows a differential loss of mRNA level in tumor stage I/II relative to that of noncancerous gastric tissue or advanced tumor stages. Through the perturbation experiments in gastric cell lines (metformin-based activation and siRNA-mediated knockdown), we demonstrated the functional relevance of AMPK α 2 loss for the AMPK signaling pathway, with downstream consequences that increase both HNF4 α and HIF-1 α . Consistent with our study, AMPK α 2 has been recently shown to suppress embryonic fibroblast transformation and tumorigenesis in animal models [41]. Thus, AMPK α 2 may represent a promising therapeutic target for gastric cancer.

Abnormal microRNA (miRNA) expression is the hallmark of many human cancer types, so one important application of RNA-seq in cancer research is to identify miRNAs that play an important role in tumorigenesis. By incorporating the expression data of protein-coding and miRNAs, we identified six key miRNAs in gastric cancer (miR-548d-3p, miR-20b, miR-135b, miR-140-3p, miR-93 and miR-19a) [40]. These key miRNAs not only show significant expression variations across different sample groups, but also have detectable repression effects on the expression of their target genes. In addition to our study, two RNA-seq studies focusing on miRNA profiling have been published. Ribeiro-dos-Santos et al. sequenced a small RNA library of normal stomach tissue and identified 15 highly expressed miRNAs [42]. Using the Solexa platform, Li et al. sequenced small RNAs from one pair of noncancerous/tumor gastric tissues [43]. Interestingly, they found that the 5p arm and 3p arm miRNAs derived from the same pre-miRNAs have different tissue preferences (noncancerous tissue vs. tumor tissue), implying a novel mechanism regulating mature miRNA selection. Although usually miRNAs are not direct therapeutic targets, identification of the key miRNAs would greatly help elucidate the molecular basis of tumor progression and aid in selecting potential therapeutic targets in gastric cancer.

5. Gene fusion detection in gastric cancer

Gene fusion refers to a hybrid gene formed from two genes as a result of translocation, interstitial deletion, or chromosomal inversion. Gene fusion can disrupt the original functions of the partner genes, generating a gene product with a new function, or it can greatly change the expression level of the partner genes (e.g., through fusion to a strong promoter). Since gene fusion can introduce dramatic functional consequences in gene function, a number of fused genes have been identified as therapeutic targets, such as EML4-ALK in non-small cell lung cancer [44,45]. Both DNA-seq and RNA-seq are able to detect gene fusion, and for that purpose, paired-end sequencing is particularly powerful.

Through whole-genome sequencing, Zang et al. found two genome rearrangements involving ERBB2, a well-known proto-oncogene and the therapeutic target of trastuzumab, in the MKN7 gastric cell line [32]. One event involves a 169 kb deletion fusing the CDK12 exon 13 to the ERBB2 intron 4; and the other event results from a 106 kb deletion fusing the NEU-ROD2 exon to the ERBB2 exon 8 (Fig. 2A and B). To our knowledge, RNA-seq has not been used to detect gene fusions in gastric cancer. However, evaluating prostate cancer using RNA-seq, Palanisamy et al. detected BRAF and RAF1 rearrangements; and then, using break-apart FISH probes, they revealed that exon 8 of the BRAF gene

was fused with exon 5 of the AGTRAP gene (angiotensin II, type I receptor-associated protein) in gastric cancer, resulting in the formation of a 597-amino acid fusion protein [46] (Fig. 2C). This result suggests that targeting RAF and MEK inhibitors may be useful in a subset of patients with this type of gene fusion. Moreover, RNA-seq can detect targetable gene fusion across different cancer types.

6. Challenges and future directions

Although the above-mentioned NGS studies provide tremendous insights into the molecular basis of gastric cancer and identify novel potential therapeutic targets, these individual lab-driven studies have limitations. First, these studies are based on small sample sizes, ranging from a handful of samples to 30 at most, which is likely due to cost and resource constraints. Because gastric cancer is a heterogeneous disease in which each cancer patient exhibits a distinct genetic and molecular profile, there is limited statistical power to accurately detect prevalent therapeutic targets based on a relatively small sample cohort. For example, only two candidate driver mutated genes (TP53 and ARID1A) were simultaneously identified by both the exome-sequencing studies [29,30] (Fig. 1). Second, DNA-seq or RNA-seq is usually employed as the only approach to characterize genomic alternations in these studies. Gastric cancer is a complex disease, involving interactions between multiple layers of aberrations. To understand how an alteration in a driver gene functions in the context of the tumor, and thereby identify and prioritize high-quality therapeutic targets, it is essential to integrate multi-dimensional genomic profiles of gastric cancer. Third, the functional data on driver genes discovered in these NGS studies are still sparse. Due to resource constraints in any individual lab, usually only one or two genes are chosen for further functional investigation. In many cases, the choice is somewhat subjective or is biased by the investigator's expertise. Ideally, a high-throughput functional assay will be used to systematically examine the functional consequences of genomic aberrations, as demonstrated by Liang et al. [47]. Fourth, the huge size of the datasets obtained through NGS make them difficult to access.

Two recently initiated consortium projects have largely overcome these four issues. One project is The Cancer Genome Atlas (TCGA), sponsored by the U.S. National Cancer Institute and the National Human Genome Research Institute. For each selected cancer type, TCGA will systematically characterize ~500 patient samples using different genomic profiling techniques, including exome sequencing, SNP arrays, copy number variation profiling, DNA methylation profiling, miRNA expression profiling, mRNA-seq, and RPPA-based protein expression. Importantly, user-friendly public access is provided for datasets generated by TCGA. The first cancer types evaluated in data released by TCGA are glioblastoma [48] and ovarian cancer [49]. Gastric cancer has been selected for the second phase. A similar effort has been undertaken by the International Cancer Genome Consortium (ICGC), with the aim of systematically studying 25,000 cancer genomes at the genomic, epigenomic, and transcriptomic levels for at least 50 cancer types. Compared with TCGA, one unique feature of the ICGC is the involvement of patients from different countries. As one of the selected cancer types, the samples of gastric cancer come from patients in both China and the United States. Over the next few years, we expect these consortium-based cancer genomic projects to be valuable resources for discovering novel therapeutic targets for gastric cancer.

Four important research areas are of particular interest for the application of NGS to gastric cancer samples. First, the incidence of gastric cancer varies greatly by ethnic group; thus, it is of interest to determine the extent to which the molecular basis of gastric cancer depends on the ethnic background of the individual (e.g.,

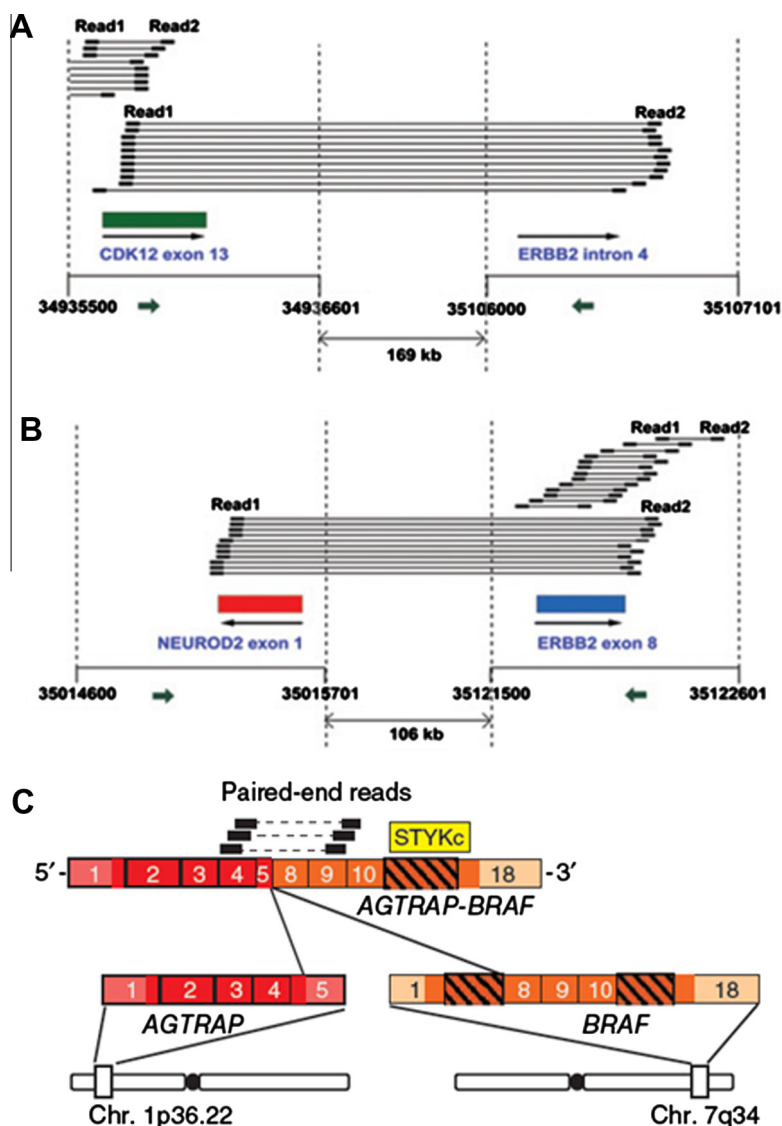


Fig. 2. Gene fusions discovered in gastric cancer by NGS studies. (A) a fusion between CDK12 exon 12 and ERBB2 intron 4 in MKN7 cells; (B) a fusion between NEUROD2 exon 1 and ERBB2 exon 8 in MKN17 cells and (C) a fusion between AGTRAP exon 5 and BRAF exon 8 in a gastric tumor sample. (A) and (B) are reproduced with permission from Zang et al. [32] and (C) is reproduced with permission from Palanisamy et al. [46].

East Asian vs. Caucasian). NGS datasets from the ICGC would be very useful to address this topic. Second, in addition to intertumoral heterogeneity (between patients with gastric cancer), intratumoral heterogeneity (within a single gastric tumor) presents an additional challenge for effective treatment, and may contribute to drug resistance. NGS can address this topic using two powerful approaches: (i) ultra-deep sequencing of the primary tumor to detect rare subclones; and (ii) low-depth sequential characterization of the tumor to identify dominant clones [50]. Third, as RNA-seq allows for the detection of splicing variants in an unprecedented manner, it is of interest to investigate the role of aberrant splicing in gastric tumorigenesis. Some unexpected targets may come from splicing variants. Fourth, using NGS to systematically identify gene fusion in gastric cancer is another approach with tremendous potential for discovering novel therapeutic targets.

Acknowledgements

This work was supported in part by grants from the U.S. National Institutes of Health: a Career Development Award through the Uterine SPORE grant (NIH/NCI P50 CA098258), TCGA GDAC

funding (NIH/NCI U24 CA143883) and the Cancer Center Support Grant to MD Anderson Cancer Center (NIH/NCI P30 CA016672). This work was also supported by UTMACC–G.S. Hogan Gastrointestinal Research Fund and the Lorraine Dell Program in Bioinformatics for Personalization of Cancer Medicine, through grants to H. Liang. The funders had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

References

- [1] F. Kamangar, G.M. Dores, W.F. Anderson, Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 24 (2006) 2137–2150.
- [2] J. Ferlay et al., GLOBOCAN 2008 cancer incidence and mortality worldwide, in: *Cancer Base, International Agency for Research on Cancer, Lyon, 2010*.
- [3] J. Volk, J. Parsonnet, Epidemiology of gastric cancer and *Helicobacter pylori*, in: T. Wang, J. Fox, A. Giraud (Eds.), *The Biology of Gastric Cancers*, Springer, 2009.
- [4] D.M. Parkin, F. Bray, J. Ferlay, P. Pisani, Global cancer statistics, *CA-A Cancer Journal for Clinicians* 55 (2005) (2002) 74–108.
- [5] M. McDonald, R.P. Hertz, S.W. Pitman Lowenthal, Pfizer facts, *The Burden of Cancer in Asia*, Pfizer, USA, 2008.

- [6] K. Yamashita, S. Sakuramoto, M. Nemoto, T. Shibata, H. Mieno, N. Katada, S. Kikuchi, M. Watanabe, Trend in gastric cancer: 35 years of surgical experience in Japan, *World Journal of Gastroenterology*: WJG 17 (2011) 3390–3397.
- [7] L. Zheng, L. Wang, J. Ajani, K. Xie, Molecular basis of gastric cancer development and progression, *Gastric Cancer: Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 7 (2004) 61–77.
- [8] C. Resende, A. Thiel, J.C. Machado, A. Ristimaki, Gastric cancer: basic aspects, *Helicobacter* 16 (Suppl. 1) (2011) 38–44.
- [9] W. Yasui, H. Yokozaki, J. Fujimoto, K. Naka, H. Kuniyasu, E. Tahara, Genetic and epigenetic alterations in multistep carcinogenesis of the stomach, *Journal of Gastroenterology* 35 (Suppl. 12) (2000) 111–115.
- [10] D. Compare, A. Rocco, G. Nardone, Risk factors in gastric cancer, *European Review for Medical and Pharmacological Sciences* 14 (2010) 302–308.
- [11] V. Bouvard, R. Baan, K. Straif, Y. Grosse, B. Secretan, F. El Ghissassi, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, V. Coglian, WIARCW, A review of human carcinogens—Part B: biological agents, *Lancet Oncology* 10 (2009) 321–322.
- [12] A.M. Machado, C. Figueiredo, R. Seruca, L.J. Rasmussen, *Helicobacter pylori* infection generates genetic instability in gastric cells, *Biochimica et Biophysica Acta* 2010 (1806) 58–65.
- [13] Y.J. Bang, E. Van Cutsem, A. Feyereislova, H.C. Chung, L. Shen, A. Sawaki, F. Lordick, A. Ohtsu, Y. Omuro, T. Satoh, G. Aprile, E. Kulikov, J. Hill, M. Lehle, J. Ruschoff, Y.K. Kang, TT Investigators, Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (TOGA): a phase 3, open-label, randomised controlled trial, *Lancet* 376 (2010) 687–697.
- [14] A.F.C. Okines, D. Cunningham, Trastuzumab in gastric cancer, *European Journal of Cancer* 46 (2010) 1949–1959.
- [15] H. Grabsch, S. Sivakumar, S. Gray, H.E. Gabbert, W. Muller, HER2 expression in gastric cancer: rare, heterogeneous and of no prognostic value—conclusions from 924 cases of two independent series, *Cellular Oncology: The Official Journal of the International Society for Cellular Oncology* 32 (2010) 57–65.
- [16] Z.A. Wainberg, A. Anghel, A.J. Desai, R. Ayala, T. Luo, B. Safran, M.S. Fejzo, J.R. Hecht, D.J. Slamon, R.S. Finn, Lapatinib, a dual EGFR and HER2 kinase inhibitor, selectively inhibits HER2-amplified human gastric cancer cells and is synergistic with trastuzumab in vitro and in vivo, *Clinical Cancer Research* 16 (2010) 1509–1519.
- [17] H.T. Arkenau, Gastric cancer in the era of molecularly targeted agents: current drug development strategies, *Journal of Cancer Research and Clinical Oncology* 135 (2009) 855–866.
- [18] G.Y. Ku, D.H. Ilson, Esophagogastric cancer: targeted agents, *Cancer Treatment Reviews* 36 (2010) 235–248.
- [19] M.A. Shah, R.K. Ramanathan, D.H. Ilson, A. Levnor, D. D'Adamo, E. O'Reilly, A. Tse, R. Trocola, L. Schwartz, M. Capanu, G.K. Schwartz, D.P. Kelsen, Multicenter phase II study of irinotecan, cisplatin, and bevacizumab in patients with metastatic gastric or gastroesophageal junction adenocarcinoma, *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 24 (2006) 5201–5206.
- [20] C. Pinto, F. Di Fabio, S. Siena, S. Cascinu, F.L. Rojas Llimpe, C. Ceccarelli, V. Mutri, L. Giannetta, S. Giaquinta, C. Funaioli, R. Berardi, C. Longobardi, E. Piana, A.A. Martoni, Phase II study of cetuximab in combination with FOLFIRI in patients with untreated advanced gastric or gastroesophageal junction adenocarcinoma (FOLCETUX study), *Annals of Oncology* 18 (2007) 510–517.
- [21] Y.F. Guan, G.R. Li, R.J. Wang, Y.T. Yi, L. Yang, D. Jiang, X.P. Zhang, Y. Peng, Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer, *Chinese Journal of Cancer* 31 (2012) 463–470.
- [22] E.R. Mardis, A decade's perspective on DNA sequencing technology, *Nature* 470 (2011) 198–203.
- [23] J. Shendure, H. Ji, Next-generation DNA sequencing, *Nature Biotechnology* 26 (2008) 1135–1145.
- [24] M. Meyerson, S. Gabriel, G. Getz, Advances in understanding cancer genomes through second-generation sequencing, *Nature Reviews Genetics* 11 (2010) 685–696.
- [25] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nature Reviews Genetics* 10 (2009) 57–63.
- [26] J.A. Schloss, How to get genomes at one ten-thousandth the cost, *Nature Biotechnology* 26 (2008) 1113–1115.
- [27] M.A. Chapman, M.S. Lawrence, J.J. Keats, K. Cibulskis, C. Sougnez, A.C. Schinzel, C.L. Harview, J.P. Brunet, G.J. Ahmann, M. Adli, K.C. Anderson, K.G. Ardlie, D. Auclair, A. Baker, P.L. Bergsagel, B.E. Bernstein, Y. Drier, R. Fonseca, S.B. Gabriel, C.C. Hofmeister, S. Jagannath, A.J. Jakubowski, A. Krishnan, J. Levy, T. Liefeld, S. Lonial, S. Mahan, B. Mfuko, S. Monti, L.M. Perkins, R. Onofrio, T.J. Pugh, S.V. Rajkumar, A.H. Ramos, D.S. Siegel, A. Sivachenko, A.K. Stewart, S. Trudel, R. Vij, D. Voet, W. Winckler, T. Zimmerman, J. Carpten, J. Trent, W.C. Hahn, L.A. Garraway, M. Meyerson, E.S. Lander, G. Getz, T.R. Golub, Initial genome sequencing and analysis of multiple myeloma, *Nature* 471 (2011) 467–472.
- [28] J.A. Stamatoyannopoulos, I. Adzhubei, R.E. Thurman, G.V. Kryukov, S.M. Mirkin, S.R. Sunyaev, Human mutation rate associated with DNA replication timing, *Nature Genetics* 41 (2009) 393–395.
- [29] W.C. Hahn, R.A. Weinberg, Rules for making human tumor cells, *The New England Journal of Medicine* 347 (2002) 1593–1603.
- [30] K. Wang, J. Kan, S.T. Yuen, S.T. Shi, K.M. Chu, S. Law, T.L. Chan, Z. Kan, A.S. Chan, W.Y. Tsui, S.P. Lee, S.L. Ho, A.K. Chan, G.H. Cheng, P.C. Roberts, P.A. Rejto, N.W. Gibson, D.J. Pocalyko, M. Mao, J. Xu, S.Y. Leung, Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer, *Nature Genetics* 43 (2011) 1219–1223.
- [31] Z.J. Zang, I. Cutcutache, S.L. Poon, S.L. Zhang, J.R. McPherson, J. Tao, V. Rajasegaran, H.L. Heng, N. Deng, A. Gan, K.H. Lim, C.K. Ong, D. Huang, S.Y. Chin, I.B. Tan, C.C. Ng, W. Yu, Y. Wu, M. Lee, J. Wu, D. Poh, W.K. Wan, S.Y. Rha, J. So, M. Salto-Tellez, K.G. Yeoh, W.K. Wong, Y.J. Zhu, P.A. Futreal, B. Pang, Y. Ruan, A.M. Hillmer, D. Bertrand, N. Nagarajan, S. Rozen, B.T. Teh, P. Tan, Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes, *Nature Genetics* 44 (2012) 570–574.
- [32] Z.J. Zang, C.K. Ong, I. Cutcutache, W. Yu, S.L. Zhang, D. Huang, L.D. Ler, K. Dykema, A. Gan, J. Tao, S. Lim, Y. Liu, P.A. Futreal, H. Grabsch, K.A. Furge, L.K. Goh, S. Rozen, B.T. Teh, P. Tan, Genetic and structural variation in the gastric cancer kinome revealed through targeted deep sequencing, *Cancer Research* 71 (2011) 29–39.
- [33] J.D. Holbrook, J.S. Parker, K.T. Gallagher, W.S. Halsey, A.M. Hughes, V.J. Weigman, P.F. Lebowitz, R. Kumar, Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine, *Journal of Translational Medicine* 9 (2011) 119.
- [34] C.F. Lee, S. Griffiths, E. Rodriguez-Suarez, A. Pierce, R.D. Unwin, E. Jaworska, C.A. Evans, S. Gaskell, A.D. Whetton, Assessment of downstream effectors of BCR/ABL protein tyrosine kinase using combined proteomic approaches, *Proteomics* 10 (2010) 3321–3342.
- [35] S.H. Amin, R. Tibes, J.E. Kim, C.P. Hybarger, Hedgehog antagonist GDC-0449 is effective in the treatment of advanced basal cell carcinoma, *The Laryngoscope* 120 (2010) 2456–2459.
- [36] S. Volinia, G.A. Calin, C.G. Liu, S. Ambs, A. Cimmino, F. Petrocca, R. Visone, M. Iorio, C. Roldo, M. Ferracin, R.L. Prueitt, N. Yanaihara, G. Lanza, A. Scarpa, A. Vecchione, M. Negrini, C.C. Harris, C.M. Croce, A microRNA expression signature of human solid tumors defines cancer gene targets, *Proceedings of the National Academy of Sciences of the United States of America* 103 (2006) 2257–2261.
- [37] T. Ueda, S. Volinia, H. Okumura, M. Shimizu, C. Taccioli, S. Rossi, H. Alder, C.G. Liu, N. Oue, W. Yasui, K. Yoshida, H. Sasaki, S. Nomura, Y. Seto, M. Kaminishi, G.A. Calin, C.M. Croce, Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis, *Lancet Oncology* 11 (2010) 136–146.
- [38] J.A. Cui, Y.B. Chen, W.C. Chou, L.K. Sun, L. Chen, J.A. Suo, Z.H. Ni, M. Zhang, X.X. Kong, L.L. Hoffman, J.S. Kang, Y.Y. Su, V. Olman, D. Johnson, D.W. Tench, I.J. Amster, R. Orlando, D. Puett, F. Li, Y. Xu, An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer, *Nucleic Acids Research* 39 (2011) 1197–1207.
- [39] A. Kahvejian, J. Quackenbush, J.F. Thompson, What would you do if you could sequence everything?, *Nature biotechnology* 26 (2008) 1125–1133
- [40] Y.H. Kim, H. Liang, X. Liu, J.S. Lee, J.Y. Cho, J.H. Cheong, H. Kim, M. Li, T.J. Downey, M.D. Dyer, Y. Sun, J. Sun, E.M. Beasley, H.C. Chung, S.H. Noh, J.N. Weinstein, C.G. Liu, G. Powis, AMPKalpha modulation in cancer progression: multilayer integrative analysis of the whole transcriptome in Asian gastric cancer, *Cancer Research* 72 (2012) 2512–2521.
- [41] K.N. Phoenix, C.V. Devarakonda, M.M. Fox, L.E. Stevens, K.P. Claffey, AMPKalpha2 suppresses murine embryonic fibroblast transformation and tumorigenesis, *Genes & Cancer* 3 (2012) 51–62.
- [42] A. Ribeiro-dos-Santos, A.S. Khayat, A. Silva, D.O. Alencar, J. Lobato, L. Luz, D.G. Pinheiro, L. Varuzza, M. Assumpcao, P. Assumpcao, S. Santos, D.L. Zanette, W.A. Silva Jr., R. Burbano, S. Darnet, Ultra-deep sequencing reveals the microRNA expression pattern of the human stomach, *PLoS ONE* 5 (2010) e13205.
- [43] S.C. Li, Y.L. Liao, M.R. Ho, K.W. Tsai, C.H. Lai, W.C. Lin, MiRNA arm selection and isomiR distribution in gastric cancer, *BMC Genomics* 13 (Suppl 1) (2012) S13.
- [44] F. Mitelman, B. Johansson, F. Mertens, The impact of translocations and gene fusions on cancer causation, *Nature Reviews Cancer* 7 (2007) 233–245.
- [45] M. Soda, Y.L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S.I. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, H. Mano, Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer, *Nature* 448 (2007) 561–U563.
- [46] N. Palanisamy, B. Ateeq, S. Kalyana-Sundaram, D. Pflueger, K. Ramnarayanan, S. Shankar, B. Han, Q. Cao, X. Cao, K. Suleman, C. Kumar-Sinha, S.M. Dhanasekaran, Y.B. Chen, R. Esgueva, S. Banerjee, C.J. LaFargue, J. Siddiqui, F. Demichelis, P. Moeller, T.A. Bismar, R. Kuefer, D.R. Fullen, T.M. Johnson, J.K. Greenson, T.J. Giordano, P. Tan, S.A. Tomlins, S. Varambally, M.A. Rubin, C.A. Maher, A.M. Chinnaiyan, Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma, *Nature Medicine* 16 (2010) 793–798.
- [47] H. Liang, L.W. Cheung, J. Li, Z. Ju, S. Yu, K. Stemke-Hale, T. Dogruluk, Y. Lu, X. Liu, C. Gu, W. Guo, S.E. Scherer, H. Carter, S.N. Westin, M.D. Dyer, R.G. Verhaak, F. Zhang, R. Karchin, C.G. Liu, K.H. Lu, R.R. Broaddus, K.L. Scott, B.T. Hennessy, G.B. Mills, Whole-exome sequencing combined with functional genomics reveals novel candidate driver cancer genes in endometrial cancer, *Genome Research* (2012) 2120–2129.
- [48] Cancer Genome Atlas Research Network, Comprehensive genomic characterization defines human glioblastoma genes and core pathways, *Nature* 455 (2008) 1061–1068.
- [49] Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma, *Nature* 474 (2011) 609–615.
- [50] F. Meric-Bernstam, G.B. Mills, Overcoming implementation challenges of personalized cancer therapy, *Nature Reviews Clinical Oncology* 9 (2012) 542–548.