Genetic and Structural Variation in the Gastric Cancer
Kinome Revealed through Targeted Deep Sequencing

Zhi Jiang Zang1,2, Choon Kiat Ong3, Ioana Cutcutache4, Willie Yu5, Shen Li Zhang2, Dachuan Huang3, Lian Dee Ler3, Karl Dykema5, Anna Gan3, Jiong Tao2,6, Siyu Lim7, Yujing Liu2,8, Bin Tean Teh3,11, and Patrick Tan1,2,12,13

University of Singapore, and Genome Institute of Singapore, Singapore Science Institute of Singapore, Yong Loo Lin School of Medicine, National NUS Graduate Medical School, 5Laboratory of Computational Biology, Medical School, 3NCCS-VARI Translational Cancer Research Laboratory, Van Andel Research Institute, Grand Rapids, Michigan; 12Cancer Trust Sanger Institute, Hinxton, and 10Section of Pathology and National Cancer Centre, 4Neuroscience and Behavioral Disorders, Duke-Centre, 2Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School Singapore, 8 College Road Singapore 169857, Singapore. E-mail : gi@duke-nus.edu.sg; or Steve Rozen, Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School Singapore, 8 College Road, Singapore 169857, Singapore. E-mail : steve.rozen@duke-nus.edu.sg; or Bin T. Teh, NCCS-VARI Translational Cancer Research Laboratory, National Cancer Centre Singapore, 11 Hospital Drive, Singapore 169610, Singapore. E-mail : bin.teh@vai.org.

doi: 10.1158/0008-5472.CAN-10-1749
©2010 American Association for Cancer Research.

Abstract

Genetic alterations in kinases have been linked to multiple human pathologies. To explore the landscape of kinase genetic variation in gastric cancer (GC), we used targeted, paired-end deep sequencing to analyze 532 protein and phosphoinositide kinases in 14 GC cell lines. We identified 10,604 single-nucleotide variants (SNV) in kinase exons including greater than 300 novel nonsynonymous SNVs. Family-wise analysis of the nonsynonymous SNVs revealed a significant enrichment in mitogen-activated protein kinase (MAPK)-related genes (P < 0.01), suggesting a preferential involvement of this kinase family in GC. A potential antioncogenic role for MAP2K4, a gene exhibiting recurrent alterations in 2 lines, was functionally supported by siRNA knockdown and overexpression studies in wild-type and MAP2K4 variant lines. The deep sequencing data also revealed novel, large-scale structural rearrangement events involving kinases including gene fusions involving CDK12 and the ERBB2 receptor tyrosine kinase in MKN7 cells. Integrating SNVs and copy number alterations, we identified Hs7467 as a cell line exhibiting both splice-site mutations and genomic amplification of MET, resulting in MET protein overexpression. When applied to primary GCs, we identified somatic mutations in 8 kinases, 4 of which were recurrently altered in both primary tumors and cell lines (MAP3K6, STK31, FER, and CDKL5). These results demonstrate that how targeted deep sequencing approaches can deliver unprecedented multilevel characterization of a medically and pharmacologically relevant gene family. The catalog of kinome genetic variants assembled here may broaden our knowledge on kinases and provide useful information on genetic alterations in GC. Cancer Res; 71(1); 29–39. ©2010 AACR.

Introduction

Protein and lipid kinases play important roles in diverse biological processes, ranging from tissue differentiation and cellular proliferation to axonal migration and organ homeostasis (1–3). As critical nodes of cellular signaling pathways, kinases frequently integrate signaling outputs of different signal transduction circuits, linking extracellular signals to nuclear gene transcription programs (1–3). In humans, genetic alterations and polymorphisms in kinases have been implicated in a wide variety of diseases, ranging from congenital defects, neurologic disease, diabetes, and cancer (4–6). On top of their medical relevance, kinases are also notable to the pharmaceutical industry as they are highly amenable to targeting by rationally designed small molecule inhibitors (7). The prominence of kinases in biology and medicine, thus, renders it essential to determine the spectrum of kinase genetic variation in healthy individuals and patients with specific diseases.

Recent advances in DNA sequencing technologies have reduced the cost of decoding human genomes by several orders of magnitude (8). At present, however, sequencing a complete human genome at sufficient depth to identify single-nucleotide variants (SNV) still represents a highly resource-intensive effort, particularly if many genomes need to be analyzed.
Recently, innovative approaches have been described for integrating such massively parallel sequencing platforms with sequence capture technologies, such as molecular inversion probes (9, 10), sequence capture arrays (11–13), and solution phase hybridization (14), enabling specific genomic regions to be selected before sequencing. Such "targeted deep sequencing" approaches are particularly suitable for rapidly profiling specific genomic regions across multiple individuals and also carry the additional advantage of being able to provide insights into other levels of variation in addition to SNVs such as structural and copy number variants.

Targeted deep sequencing surveys have been reported for various disease genes (15, 16); however, to date, targeted deep sequencing of the kinome has not been reported. Here, we applied targeted deep sequencing to characterize the extent of kinome genetic variation in gastric cancer (GC), the second leading cause of global cancer-related mortality and a disease for which kinase dysregulation has been implicated (17, 18). Surveying the protein-coding regions of 537 kinases and genes across 14 commonly used GC cell lines, we detected more than 300 novel kinase SNVs and identified mitogen-activated protein kinase (MAPK) cascade genes as frequently altered kinases in GC. We also identified structural variants and splice-site alterations affecting ERBB2 and MET, 2 pharmacologically relevant receptor tyrosine kinases (RTK). Demonstrating the applicability of this approach to primary tumors, we also discovered, for the first time in GC, somatic mutations in 7 kinase genes (MAP3K6, STK31, PAK4, FER, CDKL5, INSR, RPS6KC1). Taken collectively, our results demonstrate the power of targeted deep sequencing for generating a comprehensive catalog of molecular variants at the single-nucleotide scale, a necessary first step to exploring how such variants might affect disease susceptibility, tumorigenesis, and drug response.

Materials and Methods

GC cell lines and tissues

GC cell lines SNU5, AGS, N87, and Hs746T were obtained from the American Type Culture Collection (ATCC). Cell lines MKN1, MKN7, TMK1, IM95, AZ521, and MKN28 were obtained from the Japan Health Science Research Resource Bank (JCRB). Cell lines YCC1, YCC3, YCC11, and YCC16 were originated and obtained from Yonsei Cancer Centre (gift from Sun Yong Rha). All cell lines were tested and authenticated by the respective cell line bank (ATCC, JCRB) or originating institution (YCC) by several methods including DNA fingerprinting and/or cytogenetics. Prior to the commencement of this study, we independently reauthenticated the cell lines by comparing their genome-wide copy number (array-CGH) and mutational profiles with the published literature. The cell lines were cultured as recommended. Primary gastric tissues were obtained from the SingHealth Tissue Repository, after approvals from Institutional Research Ethics Review Committees and with signed patient informed consent.

Kinase capture arrays and deep sequencing

Kinase genomic coordinates are provided (Supplementary Table 1). Customized NimbleGen 38K sequence capture arrays were fabricated using SeqCap v2 software. Sequence capture was conducted by NimbleCap and NimbleGen linkers were ligated onto captured inserts prior to Illumina adaptors. Captured DNAs (500-bp fragments) were sequenced on the Illumina GAIIx as 76-bp paired-end reads. Image analysis and base calling were performed using the Illumina pipeline (v1.4) with default parameters.

Genome mapping and coverage computation

MAQ (Mapping and Assembly with Quality) software was used to align sequence reads to NCBI Build 36.1 reference genome (hg18; ref. 19). Sequences of NimbleGen linkers were trimmed prior to alignment. Aligned sequences were imported into R/Bioconductor (20) using the Shortread package, and read depth was calculated using the IRanges package with exon coordinates exported from BioMartV (21). Coverage of an exon was defined as the total number of aligned nucleotides within the exon coordinates divided by the length of the exon in base pairs. Hilbert curves were created using HilbertVisGUI software (22).

Detection of microgenomic genetic variants

Supplementary Figure 1 shows the discovery pipeline for novel SNVs. Uniquely aligned, different start site reads with 2 mismatches or greater to the reference genome were used to identify candidate SNVs. We retained SNVs that passed additional quality filters (Supplementary Methods). Only SNVs in exons or in canonical splice sites were analyzed. We considered an SNV to be "novel" if it was not in dbSNP (v130; ref. 23) or was in COSMIC (v44; ref. 24). UCSC databases were used for gene transcript identification and annotation of amino acid changes (25). Amino acid changes corresponding to SNVs were annotated according to the largest transcript of the gene. Nonsynonymous SNVs were submitted to PolyPhen (26) and SIFT (27) for functional prediction. MAQ was used to identify candidate microindels covering 1 or a few base pairs.

Inferring kinase structural variants

Large-scale structural variants were detected using the mapview file generated by MAQ. Only read pairs in which both reads had mapping quality greater than 30 were considered. Briefly, the mapview file was searched for clusters of ≥7 spatially clustered, same strand reads for which the matched read from the other strand was separated by ≥600 bp or on a different chromosome. For such clusters, we then examined the separation among the distant-matched reads. If the matched reads were within 500 bp of each other, we considered the cluster to represent a candidate large-scale variant. Candidates were checked for their presence in the Database of Genomic Variants (28).

Inferring copy number variations

Kinases/kinase exons with abnormally high read depths were candidates for copy number gain. These regions were detected using read depths computed by the IRanges package after importing the aligned data into R/Bioconductor. A threshold of median ± 2SD was used.

Additional procedures are described in Supplementary Materials and Methods.
Results

**Kinome capture and deep sequencing of GC cell lines**

We sequenced the kinomes of 14 GC cell lines using array-based sequence capture and Illumina GAIIx sequencing. We surveyed 537 genes, including 509 protein kinases, 23 phosphoinositide kinases, and 5 cancer-related genes (TP53, KRAS, PTEN, OGG1, and MGMT). Collectively, we targeted 8,425 exons covering 1.33 Mb of nucleotide sequence (Supplementary Table 1). For each line, we generated an average of 0.75 Gb raw sequence, using 1 flowcell lane/sample (Supplementary Table 2). We achieved a read coverage of ≥20× for 90% of the targeted exons (Fig. 1A). Hilbert plots confirmed a high degree of coverage for the majority of targeted exons (Fig. 1B), with only 76 exons (0.9%) exhibiting a coverage of greater than 5× (Supplementary Table 3), our threshold for SNV calling. Compared with high-coverage exons, exons with poor coverage tended to have either extremely high (>0.64, 90 percentile of
8,245 exons targeted) or low %GC content (<0.38, 10 percentile; $P = 1.04 \times 10^{-13}$, Supplementary Table 4), suggesting that the primary reason for the latter exhibiting poor coverage may be due to failed sequence capture (29). To assess specificity at the individual gene level, we compared exonic sequence coverage against the coverage of adjacent introns. We observed a striking accumulation of sequence reads only at the desired targeted exons and their flanking regions, but little coverage of introns within the same gene (Fig. 1C). Quantitatively, we computed a 670- to 900-fold enrichment of read coverage in targeted regions relative to nontargeted regions (Supplementary Table 5). These results confirm a sufficient degree of sequencing coverage for subsequent analysis.

**SNV identification and sequencing accuracy**

We developed a computational pipeline to discover novel SNVs (Supplementary Fig. 1). Totally, we identified 10,604 SNVs in kinase exons. Three methods were used to validate the accuracy of identified SNVs. First, we compared the SNVs with dbSNP (v130), a database of naturally occurring genetic polymorphisms (23). About 93.2% (9,882/10,604) of the SNVs were present in dbSNP (Fig. 2A). The finding that most of SNVs identified are previously observed genetic polymorphisms supports the accuracy of sequencing data.

Second, of the 722 exonic SNVs not found in dbSNP (novel SNVs), 392 correspond to synonymous SNVs, whereas the remaining 330 correspond to nonsynonymous SNVs (Fig. 2A). The finding that most of SNVs identified are previously observed genetic polymorphisms supports the accuracy of sequencing data. Second, of the 722 exonic SNVs not found in dbSNP (novel SNVs), 392 correspond to synonymous SNVs, whereas the remaining 330 correspond to nonsynonymous SNVs (Fig. 2A). We selected 234 nonsynonymous SNVs for bidirectional Sanger sequencing. For SNVs with read depth $\geq 20$, the false-positive rate was 6.1% (i.e., specificity = 93.9%). When the read depth was lowered to 5, the overall false-positive rate was 11.5%. Besides absolute read depth, we also found that "variant depth" (the number of reads where the variant is detected) is important for accuracy, as SNVs with high variant depth (>$8$) exhibited a lower false-positive rate (7%, 16/215) compared with SNVs of low variant depth (53% or 10/19). Zygosity calls based on deep sequencing were also consistent with Sanger sequencing, with 94.9% concordance.

Third, to estimate the false-negative rate (sensitivity), we compared the deep sequencing data of 2 representative cell lines (IM95 and YCC11) to genotype calls detected by Affymetrix 6.0 SNP arrays. Of 370 loci located within kinase exons that were also measurable on SNP arrays, 95% of the genotype calls were concordant with the sequencing data, indicating a false-negative rate of 5%. In addition, 9 of 9 known TP53 point mutations in our cell lines were successfully detected by deep sequencing. These results suggest that the targeted deep sequencing data are associated with a high degree of sensitivity and specificity (both >94%).

**Genetic landscape of the GC kinome**

We identified an average of 23.6 nonsynonymous and 28 synonymous novel exonic SNVs per cell line. The alteration rate varied across the lines, ranging from 97 in IM95 to 29 in MKN28 (Fig. 2B). Interestingly, N87, the second most altered cell line in our panel, is microsatellite instability (MSI) positive (30), whereas MSI-negative lines SNU5 and AGS exhibited relatively fewer SNVs (30, 31). The identification of more than 700 novel kinase SNVs indicates that a significant degree of genetic heterogeneity still awaits characterization within the GC kinome.

We focused on the 330 nonsynonymous kinase SNVs. After eliminating false-positive SNVs, there were 304 SNVs (Supplementary Table 6), including 289 missense and 15 nonsense substitutions, distributed across 194 genes (Supplementary
Table 7). Among the 304 SNVs, only 14 (4.6%) were present in COSMIC, a public database of somatically acquired mutations in cancer (24) including 9 TP53 and 2 KRAS mutations. To identify SNVs that may impact kinase function, we used PolyPhen, a computational tool for estimating the impact of amino acid substitutions incorporating information from DNA sequence, evolutionary conservation, and structural data (26). Forty-five SNVs were classified as "uncharacterized" as PolyPhen did not give a prediction score. Of the remaining 259, 60 SNVs were classified as "probably damaging" (indicating that they are likely to affect protein structure or function) and 56 were "possibly damaging" (lower level of confidence), representing 23.2% and 21.6% of the characterized SNVs, respectively (Fig. 2C). About 5.8% of the SNVs were nonsense substitutions. Taken collectively, these results suggest that one fifth of the novel SNVs found may impact on kinase function. This figure was significantly higher compared with an analogous set of kinase germline, nonsynonymous SNVs found in dbSNP (probably damaging; P-value = 7.2 × 10^{-6}; nonsense: P-value = 7.5 × 10^{-5}). To benchmark the PolyPhen predictions against an alternative program, we used SIFT, another in silico analysis tool (27), to analyze the same data (Supplementary Table 6). About 70% of the predictions were consistent between the 2 programs. Of the 45 SNVs uncharacterized by PolyPhen, SIFT offered predictions for 13 variants. Conversely, PolyPhen provided predictions for 9 of 41 variants for which SIFT could not classify. These results suggest that more functional calls can indeed be obtained by usage of additional prediction programs.

Eleven nonsynonymous SNVs were recurrently observed in the cell lines (Supplementary Table 6), including the TNK2 (R445W) SNV which was observed in 2 lines. A sequence analysis of TNK2[R445W] revealed that this alteration occurs at an evolutionarily conserved residue in the kinase domain (Fig. 2D). To determine if this SNVs might represent a germline variant, we screened for the TNK2[R445W] SNV in primary tissues. One of 48 normal gastric tissues exhibited the TNK2 (R445W) SNV indicating that it is likely a germline variant. The functional impact of this SNV on GC remains to be elucidated. Besides exonic SNVs, we also identified and verified 4 SNVs in canonical splice sites (Supplementary Table 6) and 15 putative microindels. Four of them were found in dbSNP, affecting ALK (SNU5), AURKAPS(MKN28), DMKN(MKN7), and ROCK1 (HS745T).

**Frequent alterations of MAPK signaling genes in GC**

The most frequently altered kinases were TTN (12/14 lines), followed by WNK1 (6/14), OBSCN (5/14) and SPEG (4/14; Supplementary Table 7). Notably, TTN (34,350 amino acid or aa) and OBSCN (7,968 aa) are both exceptionally large genes, which likely contributes to their being frequently mutated in human cancers (32). Of the other genes, we validated 2 WNK1 variants (I1172M, R1945C) and 1 SPEG variant (D1310N) as germline variants in primary gastric tissues (data not shown). To adjust our analysis for differences in gene size, we ranked the genes by the ratio of the number of nonsynonymous SNVs to gene size (Supplementary Table 8). TP53 and KRAS, 2 well-known oncogenes, were first and third in this ranked list whereas TTN and OBSCN fell far behind, suggesting that this approach may selectively enrich for genes with cancer-related functions.

In human cells, MAPK-related pathways play important roles in the regulation of cell proliferation, stress response, apoptosis, motility, metabolism, and DNA repair. We observed several genes related to MAPK signaling at the top of the ranked list, such as KRAS, MAP3K, and MAP2K4 (Supplementary Table 8). Indeed, many components of the MAPK cascade were altered in GC lines, especially genes in the p38, ERK, and JNK pathway (Fig. 3A). Statistical analysis confirmed that when considered as a family (56 genes, cumulative protein size 38,146 aa), MAPK signaling-related genes exhibited significantly more "non-benign" alterations (1/1,908 aa) compared with other targeted genes (480 genes, 440,170aa, 1/4,310 aa; Fisher’s exact test, corrected P-value = 0.008). The significant enrichment of MAPK-signaling genes persisted even after removing the KRAS from computation (corrected P-value = 0.04). In comparison, 3 other kinase families, protein tyrosine kinases (88 genes), NF-kB pathway kinases (48 genes), and phosphoinositide kinases (PIK3-AKT pathway kinases (31 genes) did not show similar significant enrichments of nonbenign SNVs (Supplementary Method and Table 9). Mapping the MAPK variants to the cell lines (Fig. 3B), we observed that when considered as individual genes, each MAPK-related gene exhibited alterations in only a few lines. However, when taken collectively, 11 of 14 (78%) lines exhibited "non-benign" alterations in at least 1 component of the MAPK pathway (Fig. 3B). Interestingly, the 2 lines harboring KRAS mutations (YCC16 and AGS) did not exhibit any other nonbenign MAPK gene alterations, consistent with KRAS acting as a major upstream regulator of the MAPK signaling hierarchy (Fig. 3B). These results highlight a potentially major role for perturbed MAPK signaling in GC.

**Functional consequences of MAP2K4 variants**

One MAPK-related gene recurrently altered was MAP2K4, a metastasis suppressor gene involved in multiple cancer types (33). We detected 2 novel MAP2K4 sequence alterations in the cell lines. The first was a homozygous nonsense C>T substitution at codon 121 in YCC3, leading to truncation of the MAP2K4 protein at codon 121 and disruption of the kinase domain (Fig. 4A). The second alteration, G136V, was a heterozygous G>T change in TMK1, converting an evolutionarily conserved residue in the kinase domain (Fig. 4A). PolyPhen and SIFT classified MAP2K4[G136V] as "probably damaging" and "deleterious" respectively.

To explore the biological relevance of MAP2K4 in GC, we conducted knockdown and overexpression studies (Fig. 4B). Consistent with MAP2K4 playing a possible antioncogenic role, siRNA-mediated silencing of MAP2K4 significantly enhanced the growth rate of MAP2K4 wild-type cell lines (AGS and YCC11) but had minimal effects on the proliferation rate of TMK1 cells that harbor the MAP2K4[G136V] variant (Fig. 4C). This result suggests that G136V alterations may compromise MAP2K4 protein function. In the reciprocal experiment, overexpression of wild-type MAP2K4 inhibited cell proliferation in all the lines, confirming the growth
inhibitory nature of MAP2K4 (Fig. 4D). Subsequent mutation screening of MAP2K4 in 48 gastric tumors and paired control tissue did not identify any somatic or germline alterations.

Analysis of kinase structural variants reveals ERBB2 gene fusions

Compared with conventional Sanger sequencing, 1 advantage of paired-end deep sequencing lies in its ability to identify aberrant structural variants. We examined mapped pairs of reads with inner coordinates greater than 600 bp apart (median distance = 354 bp; median absolute deviation = 40 bp). By examining such read pairs, we identified 24 potential structural variants involving kinases. Three of them (CDK11B, TYRO3, and SBK2) were found in the Database of Genomic Variants (DGV; Supplementary Table 10; ref. 28), providing confidence regarding the validity of our method. We selected 14 structural variants including 2 of the 3 in DGV for confirmatory PCR and sequencing, and validated 5 of them (35.7%). One structural variant was a 3.1 Mb deletion in chromosome X of MKN28 from positions 105,448,183 to 108,605,181, deleting GUCY2F (downstream from exon 3) and the adjacent 18 genes (Supplementary Table 10). GUCY2F catalyzes the synthesis of cGMP, a second messenger involved in multiple signaling cascades (34) and is somatically mutated in human cancer, suggesting a role for cGMP signaling in inhibiting tumorigenesis (35).

ERBB2 is a well-known proto-oncogene and the therapeutic target of trastuzumab, 1 of the first molecularly targeted therapies (36). We observed 2 genome rearrangement events involving different regions of the ERBB2 in MKN7 cells, which are known to exhibit genomic amplification of ERBB2 (37). These included a 169-kb deletion fusing CDK12 exon 13 to ERBB2 intron 4 (Fig. 5A) and a 106-kb deletion fusing NEUROD2 exon 1 to ERBB2 exon 8 (Fig. 5C). The ERBB2 structural variants were experimentally validated by Sanger sequencing and the breakpoints confirmed (Fig. 5B and D). To explore the potential expression of CDK12-ERBB2 fusion transcripts, we performed RT-PCR using primers targeted to CDK12 exon 13 and ERBB2 exon 2/3 and the adjacent 18 genes (Supplementary Table 10). CDK12 catalyzes the synthesis of cGMP, a second messenger involved in multiple signaling cascades (34) and is somatically mutated in human cancer, suggesting a role for cGMP signaling in inhibiting tumorigenesis (35).
related to 2 factors: (a) the chromosomal organization of the ERBB2 locus in MKN7 cells is likely complex due to regional genomic amplification (37) and (b) kinome resequencing does not exhaustively interrogate all possible structural arrangements in a region (e.g., rearrangements solely involving non-exonic regions will be missed). At the protein level, the CDK12–ERBB2 fusion transcripts were predicted to cause a truncated CDK12 protein but were not in-frame to ERBB2.

Western blotting confirmed the presence of a 15-kD truncated CDK12 protein in MKN7 cells, whereas ERBB2 protein expression was unaltered (Fig. 5F).

Integrative analysis of MET utilizing data from a single sequencing platform

Another intriguing aspect of targeted deep sequencing lies in the ability to integrate multiple levels of genomic information, such as SNVs, structural variants, and copy number, using data from a single sequencing platform. To investigate the association between targeted deep sequencing and copy number status, we identified kinase genes with high-read depth counts and compared these with array-CGH profiles of the same cell lines. Good overlap between the resequencing and array-CGH data were observed, as shown in Hs746T and YCC11 (Supplementary Fig. 2). Of the kinases amplified in Hs746T cells, the MET RTK is notable as MET has been shown to play an important role in cancer growth and metastasis, particularly in GC where 10% of cases have MET amplification (38). In Hs746T cells exhibiting MET genomic amplifications (Fig. 6A), we also detected a homozygous G>T substitution in MET intron13 altering the canonical splice donor site of MET exon 14 (Fig. 6B). This substitution is predicted to cause transcript skipping of exon 14, which encodes a domain important for the ubiquitination and degradation of MET protein (39). By RT-PCR, we confirmed MET transcripts in Hs746T cells are shorter due to the lack of exon 14 (Fig. 6C).

The observation that MET is both altered at the splicing and copy number level suggests that MET protein should be highly overexpressed in Hs746T. Indeed, Hs746T cells exhibited a strikingly high protein level of MET (Fig. 6D).

Kinome variation in primary gastric tumors

Finally, to demonstrate the utility of our approach in primary tissues, we analyzed 3 matched pairs of nonmalignant gastric tissues and gastric tumors. In benign gastric tissues, we detected an average of 19 nonsynonymous SNVs per tissue, and of the 304 nonsynonymous SNVs identified in cell lines, 2 SNVs were also observed in the normal gastric tissues [PIK3R2 (P4S) and PRKG1(H509Y)], indicating that these are likely to be germline variants. These results provide a measure of kinome variation in nonmalignant gastric mucosa. In the gastric
tumors, a comparison of the 3 cancer kinomes to their nonmalignant counterparts revealed potential somatic mutations in 8 kinases: MAP3K6(S291L), ATM(R337H), STK31(D115G), PAK4(E565K), FER(S250P), CDKL5(R722H), INSRR(R1022W), and RPS6KC1(R1025C). We subsequently validated these mutations by Sanger sequencing (Supplementary Fig. 3). Of these 8 kinases, 7 are novel in GC except for ATM gene mutations (40). Four of these kinases (MAP3K6, STK31, FER, and CDKL5) were also affected in at least 1 GC cell line.

Discussion

In this study, we performed a targeted deep sequencing analysis of the kinome in a panel of 14 GC cell lines. Our major goals of this study were to assess the utility of targeted sequencing for accurately identifying genetic variants, and to assemble a comprehensive catalog of kinome genetic variants in the 14 commonly used cell lines or experimental models of GC. We found the targeted deep sequencing methodology to be highly robust. Virtually, all the 8,425 targeted exons in our study were captured and we achieved an overall sequencing depth of $20 \times$ for 90% of the sequenced kinase genes. When benchmarked against Sanger sequencing and SNP genotyping arrays, we achieved an SNV detection specificity of 93.9% and sensitivity of 95%.

We identified more than 300 novel nonsynonymous SNVs in kinase exons. The novel SNVs identified likely comprise rare germline variants and somatic mutations (32). Both categories of these SNVs were validated by Sanger sequencing and PCR amplification. The Western Blot results showed consistent expression of CDK12 and ERBB2 proteins in MKN7 and control lines. Consistent with the predicted protein truncation resulting from the fusion transcripts in E, we noted a lower molecular weight CDK12 band in MKN7 under prolonged electrophoresis conditions (inset).
are important to document, as germline variants in certain kinases have been associated with an increased risk of cancer (41). In our study, a confirmed germline variant in MET was located at a conserved residue in the kinase domain, and found in 2 of 14 GC cell lines and 1 of 48 GC tissues, both of which are mostly Asian origin. It has been shown that TK2 is overexpressed in breast cancers where its expression is correlated with poor prognosis (42). It would be interesting to compare the prevalence of this novel variant in the Asian healthy and GC population.

Another interesting variant we identified was a homozygous nonsense substitution (Q37X) in the STK11/LKB1 kinase in YCC16. Inactivating germline mutations in STK11/LKB1 cause Peutz–Jeghers Syndrome (PJS), an inherited disorder associated with intestinal hamartomatous polyps and frequent gastrointestinal tumors (41). To date, 4 GC-related STK11 mutations have been reported (3757–758insT, Arg297fsX38, Leu117PhefsX46, and P324L), and of these, 3 were also associated with PJS and early onset GC (43–46). Screening for STK11/LKB1 mutations in patients with early onset GC might, thus, identify additional patients with PJS.

Our study identified several interesting findings potentially related to GC tumorigenesis. We discovered that kinases related to MAPK signaling exhibited a significantly enriched tendency to harbor "non-benign" genetic alterations, with 11 of 14 lines having at least 1 affected MAPK gene. Abnormalities in MAPK signaling have been shown to impinge on many phenotypes of cancer including independence from proliferation signals, evasion of apoptosis, and unlimited replication potential (47). One of these MAP kinases was MAP2K4, which was altered in 2 lines. Expression of MAP2K4 suppressed cell growth in vitro. Inactivating mutations in MAP2K4 have been found in approximately 5% of tumors from various tissues (33) and recent evidence supports the functional role of MAP2K4 in human cancer (48). However, to date, somatic mutations in MAP2K4 have yet to be reported in GC. Our findings suggest that it might be worthwhile to characterize MAP2K4 in an expanded panel of GC tumors.

Besides SNV detection, the kinome resequencing data allowed us to identify several kinase-related structural variants, including variants affecting GUCY2F, TYRO3, SBK2, and the ERBB2 RTK. We demonstrated the expression of novel CDK12-ERBB2 fusion transcripts. However, as these transcripts are not expected to disrupt ERBB2 protein translation, their functional importance remains to be determined. Finally, the ability to infer genetic variation at multiple levels (SNVs, structural variants, and copy number) allowed us to perform integrative analysis utilizing data from a single deep sequencing platform. Using the example of the MET proto-oncogene, we identified Hs746T where MET was both amplified and associated with a splice-site point mutation. Such "dual mechanisms" of oncogene mutation and amplification have also been proposed for EGFR in lung cancer (49). Interestingly, in lung cancer, activation of MET by gene amplification and by splice-site mutations deleting the juxtamembrane domain appears to be mutually exclusive (50), suggesting that either type of the alterations may confer growth advantage to lung epithelial cells. In contrast, our results reveal that in GC, a paradigm of "dual mechanisms" (amplification and activating mutation) may impinge on MET.

In conclusion, we identified more than 300 novel micro- and macrogenomic alterations involving kinases, many of which may contribute to GC development. MAPK-signaling genes are identified as frequently altered kinases in our study, suggesting a causal role for dysregulated MAPK pathway in GC tumorigenesis. Our results may contribute to understanding the genetic architecture of this important subset of the human
genome in GC and facilitate the usage of these GC models in the laboratory.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Kalpana Ramnarayan, Minghui Lee, and Jeanie Wu. We also acknowledge the generosity of the Lee Foundation who contributed to the establishment of the Lee Foundation Genome Sequencing Facility at NCCS.

Grant Support

This project was funded by grants to P. Tan from A-star, NMRC, Duke-NUS and CSIS. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. Received May 18, 2010; revised September 22, 2010; accepted October 14, 2010; published online November 19, 2010.

References


Genetic and Structural Variation in the Gastric Cancer Kinome Revealed through Targeted Deep Sequencing

Zhi Jiang Zang, Choon Kiat Ong, Ioana Cutcutache, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-10-1749

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2010/11/18/0008-5472.CAN-10-1749.DC1

Cited articles
This article cites 50 articles, 14 of which you can access for free at:
http://cancerres.aacrjournals.org/content/71/1/29.full#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/71/1/29.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.