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

Supplementary Materials and Methods
Supplementary Result
Legends to Supplementary Tables 1-11
Legends to Supplementary Figure 1- 3

Zhi Jiang Zang[1, 3], Choon Kiat Ong[2], Ioana Cutcutache[4], Willie Yu[2], Shen Li Zhang[3], Dachuan Huang[2], Lian Dee Ler[2], Karl Dykema[5], Anna Gan[2], Jiong Tao[3,6], Siyu Lim[8], Yujing Liu[3,7], P. Andrew Futreal[9], Heike Grabsch[10], Kyle A. Furge[5], Liang Kee Goh[3], Steve Rozen[4]*, Bin Tean Teh[2,11]*, Patrick Tan[1,3,12,13]*

[1]Cellular and Molecular Research/[2]NCCS-VARI Translational Cancer Research Laboratory, National Cancer Centre, 11 Hospital Drive, Singapore, 169610
[3]Cancer and Stem Cell Biology Program/[4]Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, 8 College Road, Singapore, 169857
[5]Laboratory of Computational Biology/[11]Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA
[6]Department of Physiology/[7]Singapore-MIT Alliance/[12]Cancer Science Institute of Singapore, 4 Engineering Drive 3, Yong Loo Lin School of Medicine, National University of Singapore, 5 Lower Kent Ridge Road, Singapore, 119074
[8]Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798
[9]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK
[10]Section of Pathology and Tumour Biology, Leeds Institute of Molecular Medicine, St James's University Hospital, Beckett Street, Leeds, LS9 7TF. UK
[13]Genome Institute of Singapore, 60 Biopolis Street Genome #02-01, Singapore, 138672

Running Title: Targeted Sequencing of the Gastric Cancer Kinome

Key Words: Kinase; Massively parallel sequencing; Gastric cancer

*Address correspondence to steve.rozen@duke-nus.edu.sg, or bin.teh@vai.org, or gmstanp@duke-nus.edu.sg

Supplementary Materials and Methods

DNA and RNA extraction

Genomic DNA was extracted from cell lines using a Qiagen genomic DNA extraction kit. Total RNA was extracted from cells using Trizol (Invitrogen, CA), digested with RNase free DNase (RQ1 DNase, Promega) at 37^{o}C for 30 minutes, and subsequently purified using the RNeasy Mini kit (Qiagen,CA).

Sanger sequencing validation

Primers for sequencing validation were designed using the Primer3 (http://frodo.wi.mit.ed/primer3/). Purified PCR products were sequenced in forward and reverse directions using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Version 3) and an ABI PRISM 3730 Genetic Analyzer (Applied Biosystems, CA). Chromatograms were analyzed by SeqScape V2.5 and manual review. The validation PCR primers are listed in Supplementary Table 11. To validate candidate large-scale variants, we used PCR to amplify across putative breakpoints (using genomic DNA as the template), followed by Sanger sequencing of the PCR products. Supplementary Table 11 lists the PCR primers used to validate the candidate large-scale variants.

Copy number and SNP microarrays

For copy number, gastric cancer cell line genomic DNAs and reference DNAs were labeled and hybridized to Agilent human genome 244K CGH microarrays according to

the manufacturer's instructions. Ratios were plotted according to genomic order using R/Bioconductor. For SNP calls, genomic DNAs were hybridized onto Affymetrix SNP Array 6.0 microarrays according to the manufacturer's instructions. SNPs from the Affymetrix Genome-Wide Human SNP 6.0 overlapping with kinome exon regions were filtered based on a call rate threshold of > 95%.

Cell line transfections

Recombinant DNA constructs (OriGen, MD) or siRNAs (Santa Cruz, CA) were transfected into gastric cancer cell lines using Lipofectamine (Invitrogen, CA) according to the manufacturer's conditions.

Real-time PCR and RT-PCR

Total RNAs were reverse transcribed to cDNA using iScript™ cDNA Synthesis Kit (Bio-Rad, CA) following the manufacturer's instructions. *MAP2K4* and *GAPDH* real-time PCR were performed with SsoFast EvaGreen Supermix according to manufacturer's recommendations using a CFX96™ Real-Time PCR Detection System (Bio-Rad, CA) and the following primers: *MAP2K4*-F (AATTCGGTCAACAGTGGATG); *MAP2K4*-R(TCCAACAGTCACCCTCTCTG); *GAPDH*-F(AAGGTCATCCCTGAGCTGAA);*GAPDH*-R(TGACAAAGTGGTCGTTGAGG). *MET* RT-PCR was performed using forward (CACAGGATTGATTGCTGGTGTT) and reverse (ATGCACAATCAGGCTACTGG) primers. *CDK-12/ERBB2* RT-PCR was performed using forward (TGGCAGTTCTCTTGAGTCAGC) and reverse (CAGGTAGGTGAGTTCCAGGTTTCC) primers.

Western blot analysis

Cellular proteins were extracted with PBS containing 0.1% (v/v) Triton X-100 in the presence of protease inhibitors. Proteins resolved by SDS-PAGE were electroblotted to a nitrocellulose membrane (Amersham, Buckinghamshire) and the membrane was incubated overnight at 4°C with blocking buffer (PBS containing 5% (w/v) skim milk and 0.05% (v/v) Tween-20). Primary and secondary antibody incubations were done in blocking buffer. Anti-MAP2K4 (sc-837), anti-CDK12 (sc-81834) and anti-ERBB2 (sc-52349) antibodies were purchased from Santa Cruz Biotechnology (CA) and anti-α-Tubulin antibody was from Sigma (LA). The membranes were washed with PBS containing 0.05% (v/v) Tween-20 followed by analysis using the Supersignal Chemiluminescent kit (Pierce, IL) according to the manufacturer's recommendations.

Cell proliferation assays

Cells were seeded into 96-well plates (2×10^3 cells per well) 24 hours prior to transfection with recombinant DNA constructs or siRNAs. Growth of untransfected or 24, 48 and 72 hours post-transfected cells were determined by the colorimetric 3-(4,5-dimethylthiazol-2yl)-5-(3-carboxymethoxyphenyl)-(4-sulfophenyl)-2H-tetrazoluim assay according to the manufacturer's instruction (MTS; Promega, WI). Experiments were performed in triplicate and repeated at least twice.

Statistical analysis

Chi-square tests were used to compute proportions of novel exonic nonsynonymous SNVs in different functional categories to dbSNP-reported exonic nonsynonymous SNVs.

Fisher's exact tests were used to compare the proportions of "non-benign" SNVs in MAP kinase (MAPK) pathway genes versus the other on-targeted genes (excluding *TP53)*. In the MAPK analysis, we considered four different kinase families, including MAPK-related genes, PTKs (protein tyrosine kinases), NFκB(nuclear factor kappa-light-chain-enhancer of activated B cells) pathway kinases and *PIK3* (Phosphoinositide 3-kinase)-*AKT* pathway genes. All four of these kinase families were comparably represented across the kinase population, with 56 MAPK genes (cumulative protein size of 38146 aa), 88 *PTK*s (84468 aa), 48 *NFκB*-kinases (35168 aa), and 31 *PI3K-AKT* genes (26575 aa) respectively. P-values were adjusted for multiple hypotheses using a Bonferroni correction (4 tests), with a p-value of $p<0.05$ being deemed significant (Supplementary Table 9).

Supplementary Results:

Sequence (forward) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 500bp product:
GGAAACAACAGTGACAAGAACAGTGGGCCACAGGGGCCCCGAAGAACTCCC
ACAATGCCACAGGAGGAGGCAGCAGTCGCAWTTGARGTACCACCTCCCGAG
GGTGATTGCTTCCCCATGCGGGGTAGAACCTTTGCTGTCCTGTTCACCACTCT
ACCTCCWGCACAGAATTTGGCTTATGCCTACTCAATGTGAAGATGATGAGGA
TGAAAACCTTTGTGATGATCCACTTCCACTTAATGAATGGTGGCAAAGCAAA
GCTATATTCAAGACCACATGCAAAGCTACTCCCTGAGCAAAGAGTCACAGAT
AAAACGGGGGCACCAGTAGAATGGCCRSAACAAACGCAGTGCAKCACAGAG
ACTCARACCCTGGCAGCCATGCCTGCGCAGGCAGTGATGAGAGTGACATGTA
CTGTTGTGGACATGCACAAAA

Sequence (reverse) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 500bp product:
CAGGCATGGCTGCCAGGGTCTGAGTCTCTGTGCTGCACTGCGTTTGTCCTGGC
CATTCTACTGGTGCCCCCGTTTTATCTGTGACTCTTTGCTCAGGGAGTAGCTTT
GCATGTGGTCTTGAATATAGCTTTGCTTTGCCACCATTCATTAAGTGGAAGTG
GATCATCACAAAGGTTTTCATCCTCATCATCTTCACATTGAGTAGGCATAAGC
CAAATTCTGTGCTGGAGGTAGAGTGGTGAACAGGACAGCAAAGGTTCTACCC
CGCATGGGGAAGCAATCACCCTCGGGAGGTGGTACTTCAATTGCGACTGCTG
CCTCCTCCTGTGGCATTGTGGGAGTTCTTCGGGGCCCCTGTGGCCCACTGTTC
TTGTCACTGTTGTTTTCCTCCAGACTGCCTGCTGGCTCTTGGGTTTTCATCAGC
TGACTCAAGAGAACTGCCAA

Sequence (forward) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 400bp product:
AGGAAACAACAGTGACAAGAACAGTGGGCCACAGGGGCCCCGAAGAACTYC
CACAATGCCACAGGAGGAGGCAGCAGCCTACTCAATGTGAAGATGATGAGG
ATGAAAACCTTTGTGATGATCCACTTCCACTTAATGAATGGTGGCAAAGCAA
AGCTATATTCAAGACCACATGCAAAGCTACTCCCTGAGCAAAGAGTCACAGA
TAAAACGGGGGCACCAGTAGAATGGCCAGGACAAACGCAGTGCAGCACAGA
GACTCAGACCCTGGCAGCCATGCCTGCGCAGGCAGTGATGAGAGTGACATGT
ACTGTTGTGGACATGCACAAAAGTGAGTGTGCATGTMCAC

Sequence (reverse) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 400bp product:
GCAGGCATGGCTGCCAGGGTCTGAGTCTCTGTGCTGCACTGCGTTTGTCCTGG
CCATTCTACTGGTGCCCCCGTTTTATCTGTGACTCTTTGCTCAGGGAGTAGCTT
TGCATGTGGTCTTGAATATAGCTTTGCTTTGCCACCATTCATTAAGTGGAAGT
GGATCATCACAAAGGTTTTCATCCTCATCATCTTCACATTGAGTAGGCTGCTG
CCTCCTCCTGTGGCATTGTGGGAGTTCTTCGGGGCCCCTGTGGCCCACTGTTC
TTGTCACTGTTGTTTTCCTCCAGACTGCCTGCTGGCTCTTGGGTTTTCATCAGC
TGACTCAAGGAACTGCCAA

Sequence (forward) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 800bp product:
CAACAGTGACAAGAACAGTGGGCCACAGGGGCCCCGAASAACTCCCACAAT
GCCACAGGAGGAGGCAGCAGTCGCAATTGAAKTACCACCTCCCGAGGGTGAT

TGCTTCCCCWTGCGGGGTAGAACCTTTGCTGTCCTGTTCACCACTCTACCTCC
AGCACAGAATTTGGCTTATGCCTACTCAATGTGAAGATGATGAGGATGAAAA
CCTTTGTGATGATCCACTTCCACTTAATGAATGGTKCCTCRTCTCCTCTTCCTT
GGACCAAAAGTCARAGGCWASAATGYCCTCATTCATACCCCAATGGTCTATA
CCTCCMGCARCAAGGCGAGTGAGCAAGTGATGTCCTGAAAGGCCCAGTGGA
TCAGTGGAATGAAGCGGGCAGGAAGACTTAATGCTCCTGAAACAAGGAATCC
AGAATCCARGARAAGGATGGCTAAGTGGGGCTTTCAAGGKACAAGTATGGG
GGTTGAAGGRGTCRCTGTCCCTATACCAAATCCGAAAATATTGTGACCAGGA
ACCATTCTGTCCMACTCTTCTATTTCAGGTGGCAAAGMAAAGCTATATTCAA
GACCACATGCAAAGCTACTCCCTGAGCAAAGAGTCRCAGATAATACRGGGGC
ACCAGTAGAATGGCCAGGACAAACGCAGTGCAGCACAGAGACTCAGACCCT
GGCAGCCATGCCTGCGCAGGCAGTGATGAGAGTGACATGTACTGTTGTGGAC
ATGCAC

Sequence (reverse) of RT-PCR (CDK12 exon13 and ERBB2 exon5) 800bp product:
CGCAGGCATGGCTGCCAGGGTCTGAGTCTCTGTGCTGCACTGCSTTTGTCCTG
GCCATTCTACTGGTGCCCCCGTTTTATCTGTGACTCTTTGCTCAGGGAGTAGC
TTTGCATGTGGTCTTGAATATAGCTTTGCTTTGCCACCTKWMATWRAAKAGW
WGGACAKAATGGTTCCTGGTCACAATATTTTCGGATTTGGTATARGGACAGT
GACCCCTTCAACCCCCATACTTGTCCCTTGAAARMCCCRCTGASCCATCCTTC
TCCTGGATTCTGGATTCCTTGTTTCASGAGMACTAAKTCTTCCTGCCCGCTTC
ATTCCWCTGATCCACTGGGCCTTTCAGGACATCACTTGCTCACTCGACTTGCT
GCTGGAGGTATAGACCACTGGGGTATGAATGAGGGCATTCTTGCCTCTGACT
TTTGKKMCAAGGAAGARSAGACRAGGAACCATTCATTAAGTGGAAGTGGATC
ATCACAAAGGTTTTCATCCTCATCATCTTCACATTGAGTAGGCATAAGCCAAA
TTCTGTGCTGGAGGTAGAGTGGTGAACAGGACAGCAAAGGTTCTACCCCGCA
TGGGGAAGCAATCACCCTCRGGAGGTGGTACTTCAATTGCGACTGCTGCCTC
CTCCTGTGGCATTGTGGGAGTTCTTCGGGGCCCCTGTGGCCCACTGTTCTTGT
CACTGTTGTTTTCCTCCAGACTGCCTGCTGGCTC

Legends to Supplementary Tables

Supplementary Table 1

List of genes and exons targeted for sequence capture and Illumina GAIIx deep sequencing.

Supplementary Table 2

Illumina GAIIx kinome sequencing throughput and filtering summary of 14 gastric cancer cell lines.

Fourteen gastric cell lines, each occupying one flowcell lane, are listed according to the order of sequencing. **Kinase Genes bps** are defined as sequenced bases that are within either exons or introns of kinase genes. **Kinase Exon bps** are defined as sequenced bases only within the exons of kinase genes. **Column 2** represents the total number of sequenced bases that can be aligned to the reference genome build36.1. **Column 3** shows the number of sequenced bases remaining after removing PCR duplicates, accepting reads with less than or equal to two mismatches, and reads with a unique alignment to the reference genome build36.1. **Column 4** shows the number of sequenced bases falling within the kinase gene exons or introns. **Column 5** shows the filtered sequenced bases falling within the exons or introns of kinase genes after removing reads due to PCR duplication, having greater than two mismatches or not uniquely aligning to the reference genome. **Column 6** shows the filtered sequenced bases falling only within the exons of kinase genes after removing reads due to PCR duplication, having greater than two mismatches or not uniquely aligning to the reference genome.

Supplementary Table 3

List of targeted exons receiving low coverage across all samples (read depth< 5).

Supplementary Table 4

Differences in GC content in targeted exons showing high or low read coverage.

The last column represents exons showing extreme deviations in GC content (top 10% or bottom 10%), with absolute values being GC<0.38 (lowest 10%) or GC>0.64 (highest 10%). A read depth of ≥5 was used as the threshold for variant calling in the Results. The last row (Proportion) represents the fraction of low coverage exons divided by the total number of exons in the column. There is a significant enrichment in the proportion of low coverage exons exhibiting extreme GC content ($p=1.04\times10^{-13}$, Fisher's exact test).

Supplementary Table 5

Quantitative comparison of read coverage between targeted and non-targeted regions of the 14 lines sequenced.

"Non-targeted bps" refers to the number of base pairs associated with genomic regions other than targeted kinase exons. "Non-targeted coverage" refers to "Non-targeted bps" divided by "the size of human genome – size of the targeted kinome". "Kinase exons bps" refers to the number of base pairs associated with targeted kinase exons. "Kinase coverage" refers to "Kinase exons bps" divided by "the size of targeted kinome". "Enrichment" refers to "Kinase coverage" divided by "Non-targeted coverage".

Supplementary Table 6

List of novel exonic nonsynonymous SNVs and splice site variants found in 14 gastric cancer cell lines.

The SNV location, gene name, cell line, corresponding amino acid alteration, Sanger sequencing validation status, total read depth, variant depth, PolyPhen and SIFT functional annotation are listed.

Supplementary Table 7

The most altered genes sequenced in the 14 gastric cancer cell lines

Supplementary Table 8

Genes with novel exonic nonsynonymous variants found in our study, ranked by the ratio of the number of nonsynonymous variants in the gene and the largest transcript size of the gene.

Supplementary Table 9

Enrichment of non-benign SNVs in the MAPK cascade genes, but not in PIK3-AKT, NF-κB , or PTK family genes.

Supplementary Table 10

Kinase structural variants inferred from paired-end deep sequencing data.

[a] Region 1 = the region that contains a kinase. In cases where both regions (or none of them) contain kinases, then the ordering is based on genomic position; [b] Paired Read

Depth refers to the number of paired reads that support the variant; [c] DGV = Database of Genomic Variants (http://projects.tcag.ca/variation/); [d] NA = Not applicable; DGV does not record inter-chromosomal translocations; [e] NT = Not tested

Supplementary Table 11

Primers used for Sanger sequencing of novel exonic nonsynonymous SNVs, splice site variants and structural variants.

Legend to Supplementary Figures

Supplementary Figure 1. The SNV discovery pipeline used in this study.

Supplementary Figure 2. Copy number alterations of kinases revealed by targeted deep sequencing and comparison to aCGH.

Chromosomal plots (1-22) showing kinase read depths by targeted deep sequencing and copy number alteration by aCGH (Agilent 244K) of Hs746T and YCC11. Kinases with high read depth are labeled and are largely concordant with aCGH data.

Supplementary Figure 3. Chromotogram of kinase somatic mutations identified from gastric tumors and matched non-malignant controls using Sanger sequencing.