

# UC San Diego

## UC San Diego Previously Published Works

### Title

Mutational landscape of gastric adenocarcinoma in Chinese: Implications for prognosis and therapy

### Permalink

<https://escholarship.org/uc/item/52p5x2d1>

### Journal

Proceedings of the National Academy of Sciences of the United States of America, 112(4)

### ISSN

0027-8424

### Authors

Chen, Kexin  
Yang, Da  
Li, Xiangchun  
et al.

### Publication Date

2015-01-27

### DOI

10.1073/pnas.1422640112

Peer reviewed

# Mutational landscape of gastric adenocarcinoma in Chinese: Implications for prognosis and therapy

Kexin Chen<sup>a,1</sup>, Da Yang<sup>b,c,1</sup>, Xiangchun Li<sup>d,e,1</sup>, Baocun Sun<sup>f</sup>, Fengju Song<sup>a</sup>, Wenfeng Cao<sup>f</sup>, Daniel J. Brat<sup>g</sup>, Zhibo Gao<sup>d</sup>, Haixin Li<sup>h</sup>, Han Liang<sup>h</sup>, Yanrui Zhao<sup>a</sup>, Hong Zheng<sup>a</sup>, Miao Li<sup>d</sup>, Jan Buckner<sup>i</sup>, Scott D. Patterson<sup>j</sup>, Xiang Ye<sup>k</sup>, Christoph Reinhard<sup>l</sup>, Anahita Bhatena<sup>m</sup>, Deepa Joshi<sup>n</sup>, Paul S. Mischel<sup>o</sup>, Carlo M. Croce<sup>p</sup>, Yi Michael Wang<sup>q</sup>, Sreekumar Raghavakaimal<sup>q</sup>, Hui Li<sup>a,f,h</sup>, Xin Lu<sup>m</sup>, Yang Pan<sup>j</sup>, Han Chang<sup>r</sup>, Sujuan Ba<sup>q,5</sup>, Longhai Luo<sup>d,t</sup>, Webster K. Cavenee<sup>o,2</sup>, Wei Zhang<sup>b,2</sup>, and Xishan Hao<sup>a,f,h,2</sup>

<sup>a</sup>Department of Epidemiology and Biostatistics, <sup>f</sup>Department of Pathology, and <sup>h</sup>Department of Gastrointestinal Surgery, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, People's Republic of China; <sup>b</sup>Department of Pathology, University of Texas MD Anderson Cancer Center Informatics Center, Houston, TX 77030; <sup>c</sup>Center of Pharmacogenetics, Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15261; <sup>d</sup>Beijing Genomics Institute-Shenzhen, Shenzhen 518083, Guangdong, People's Republic of China; <sup>e</sup>Department of Medicine and Therapeutics, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong 999077, People's Republic of China; <sup>g</sup>Department of Oncology Discovery, Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322; <sup>i</sup>Division of Medical Oncology, Mayo Clinic, Rochester, MN 55905; <sup>j</sup>Medical Sciences, Amgen Inc., Thousand Oaks, CA 91320; <sup>k</sup>Lilly China Research and Development Center and <sup>l</sup>Translational Sciences Oncology, Eli Lilly and Company, Indianapolis, IN 46285; <sup>m</sup>Translational Oncology & Personalized Medicine, AbbVie Inc., North Chicago, IL 60064; <sup>n</sup>Exploratory Clinical & Translational Research, Bristol-Myers Squibb India, Mumbai 400013, India; <sup>o</sup>Applied Genomics and Bioinformatics, Bristol-Myers Squibb, Princeton, NJ 08540; <sup>p</sup>Ludwig Institute for Cancer Research, University of California at San Diego, La Jolla, CA 92093; <sup>q</sup>Human Cancer Genetics Program, Ohio State University, Columbus, OH 43210; <sup>r</sup>The National Foundation for Cancer Research, Bethesda, MD 20814; <sup>s</sup>Asian Fund for Cancer Research, Hong Kong, People's Republic of China; and <sup>t</sup>Department of Biology, University of Copenhagen, DK-1165 Copenhagen, Denmark

Contributed by Webster K. Cavenee, December 15, 2014 (sent for review November 20, 2014)

**Gastric cancer (GC) is a highly heterogeneous disease. To identify potential clinically actionable therapeutic targets that may inform individualized treatment strategies, we performed whole-exome sequencing on 78 GCs of differing histologies and anatomic locations, as well as whole-genome sequencing on two GC cases, each with three primary tumors and two matching lymph node metastases. The data showed two distinct GC subtypes with either high-clonality (HiC) or low-clonality (LoC). The HiC subtype of intratumoral heterogeneity was associated with older age, *TP53* (tumor protein P53) mutation, enriched C > G transition, and significantly shorter survival, whereas the LoC subtype was associated with younger age, *ARID1A* (AT rich interactive domain 1A) mutation, and significantly longer survival. Phylogenetic tree analysis of whole-genome sequencing data from multiple samples of two patients supported the clonal evolution of GC metastasis and revealed the accumulation of genetic defects that necessitate combination therapeutics. The most recurrently mutated genes, which were validated in a separate cohort of 216 cases by targeted sequencing, were members of the homologous recombination DNA repair, Wnt, and PI3K-ERBB pathways. Notably, the drugable *NRG1* (neuregulin-1) and *ERBB4* (V-Erb-B2 avian erythroblastic leukemia viral oncogene homolog 4) ligand-receptor pair were mutated in 10% of GC cases. Mutations of the *BRCA2* (breast cancer 2, early onset) gene, found in 8% of our cohort and validated in The Cancer Genome Atlas GC cohort, were associated with significantly longer survivals. These data define distinct clinicogenetic forms of GC in the Chinese population that are characterized by specific mutation sets that can be investigated for efficacy of single and combination therapies.**

clonality | exome sequencing | mutation | ERBB | BRCA2

**G**astric cancer (GC) is the fourth most common cancer and the second leading cause of cancer death worldwide, accounting for 8% of all newly diagnosed cancers and 10% of cancer mortality(1). Environmental risk factors for GC include a high-salt diet, smoking, and infectious agents (1), including the bacterium *Helicobacter pylori* (2), and Epstein Barr Virus (3). Consistent with its complicated etiology (e.g., diet) and anatomical environment, GC is clinically and pathologically highly heterogeneous (4), with a large variation in 5-y survival rates in different countries, and even different cities in the same country (5, 6). This clinical heterogeneity is mirrored by concomitant heterogeneous molecular signatures in GC mRNA, protein, and

miRNA expression profiles (7, 8). Standard treatment strategies have largely ignored the heterogeneity and individuality of different subtypes of GC. The current approach entails surgical removal of the tumor followed by adjuvant fluoropyrimidine, taxane, and platinum-based chemotherapy doublets or triplets, especially for advanced GC, and this is exacerbated by the lack of reliable markers to predict response. Recently, the US Food and Drug Administration and the European Medicines Agency have approved Trastuzumab for patients with HER2-overexpressing metastatic GC, which represent less than 15% of the disease population. The high incidence of GC in Asian countries and its increasing incidence in Western countries point to a clear need for developing more effective therapies for GC as well as the discovery of markers that predict their therapeutic response.

Genome sequencing has emerged as a powerful tool to identify potential driving oncogenic targets for therapeutic intervention. Wang et al. sequenced 22 samples from Hong Kong GC patients and identified mutations in genes involved in chromatin modification

## Significance

**We have identified a lethal subtype of gastric cancer (GC) that is characterized by high levels of clonal heterogeneity and *TP53* (tumor protein P53) mutation. We have also uncovered key novel mutations in the targetable *NRG1* (neuregulin-1) and *ERBB4* (V-Erb-B2 avian erythroblastic leukemia viral oncogene homolog 4) ligand-receptor pair and identified *BRCA2* (breast cancer 2, early onset) mutations as new genetic markers to predict better survival for GC. Our study represents a novel approach for GC personalized medicine and identified novel clinical actionable therapies for GC therapy.**

Author contributions: K.C., B.S., W.C., D.J.B., J.B., S.D.P., X.Y., C.R., A.B., D.J., P.S.M., C.M.C., Y.M.W., S.K., Y.P., H.C., S.B., W.K.C., W.Z., and X.H. designed research; Z.G., H. Liang, M.L., X. Lu, and L.L. performed research; D.Y., Haixin Li, Y.Z., H.Z., Hui Li, X. Lu, and W.Z. analyzed data; and K.C., D.Y., X. Li, F.S., W.C., X. Lu, W.K.C., and W.Z. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequence reported in this paper has been deposited in the European Genome-phenome Archive database, [www.ebi.ac.uk](http://www.ebi.ac.uk) (accession no. EGAS00001001056).

<sup>1</sup>K.C., D.Y., and X. Li contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [wcavenee@ucsd.edu](mailto:wcavenee@ucsd.edu), [xishanhao@sina.com](mailto:xishanhao@sina.com), or [wzhang@mdanderson.org](mailto:wzhang@mdanderson.org).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422640112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1422640112/-DCSupplemental).

(e.g., *ARID1A*) (9). Two recent and more comprehensive genomic characterizations of Russian and Hong Kong GC by the The Cancer Genome Atlas (TCGA) and Wang et al. (10, 11), respectively, reported anatomical and histological subtype-specific genetic and epigenetic alterations. Both studies highlighted *RHOA* (ras homolog family member A) mutation as a novel oncogenic driver for the diffuse subtype of GCs, which was functionally validated by Wang et al. (11) and Kakiuchi et al. (12). Hotspot mutations in *RHOA* were shown to result in defective RHOA signaling and promote escape from anoikis in organoid cultures. These studies revealed new insights into the genetic and epigenetic alterations in GC, but their lack of sufficient clinical follow-up data limited their potential translation into clinical utility.

Our international consortium sought to determine whether novel or clinically actionable events for Chinese GC could be identified by focusing on its heterogeneity (anatomical location, known histologic subtypes, sex) and analyzing samples with comprehensive clinical follow-up. This strategy allows us to report herein several clinically important findings. First, we identified a new lethal subtype of GC that is characterized by high levels of clonal heterogeneity. Second, through whole-genome sequencing (WGS) of multiple samples from two patients, we found that clonal divergence occurs intratumorally and mutations accumulate during clonal metastasis. Third, we uncovered key novel mutations in the targetable *neuregulin-1* (*NRG1*) and *ERBB4* ligand-receptor pair in 10% of GC cases. Finally, we show that *BRCA2* mutations serve as new genetic markers to predict longer survival for GC.

## Results

Frozen tissue samples of primary gastric adenocarcinoma from 294 northern Chinese patients not treated with chemotherapy or radiotherapy prior were derived from surgical resection specimens and obtained from the Tianjin Medical University Cancer Institute and Hospital-National Foundation for Cancer Research Joint Tissue Banking Facility (Dataset S1). Germ-line DNA was available from matching blood samples and used as a sequence reference to detect somatic alterations. All patients signed informed consent, and the study protocol was approved by the Institutional Review Board from Tianjin Medical University Cancer Institute and Hospital. We performed whole-exome sequencing (WES) on 78 samples (discovery cohort) and WGS on two cases with three primary tumors and two matching lymph node metastases. Novel recurrently mutated genes were further validated in a separate cohort of 216 cases from the same population (validation cohort) by targeted sequencing using the Ion Torrent PGM platform. Primary tumors were from the following anatomic locations within the stomach: 64 from the antrum (21.77%), 118 from the body (40.14%), and 112 from the cardia (38.10%). All histopathologic diagnoses were subjected to independent review by at least two senior pathologists, and in this study cases with discrepant diagnosis among pathologists (five cases, 1.70%) were not included. In this study, 124 cases were classified as intestinal-type, 152 as diffuse-type, and 18 of mixed intestinal and diffuse histology. Clinical stage was determined according to National Comprehensive Cancer Network guidelines. Among the 294 samples, 6 (2.05%) were stage I, 85 (28.91%) were stage II, 97 (32.99%) were stage III, and 106 (36.05%) were stage IV. At the time of this report, clinical follow-up data were complete for 284 participants, with 13.84 months of median follow-up, 25.08 months for the WES series (32 deceased, 41.03%), and 12.16 months for those with targeted sequencing (61 deceased, 28.24%) (Dataset S1).

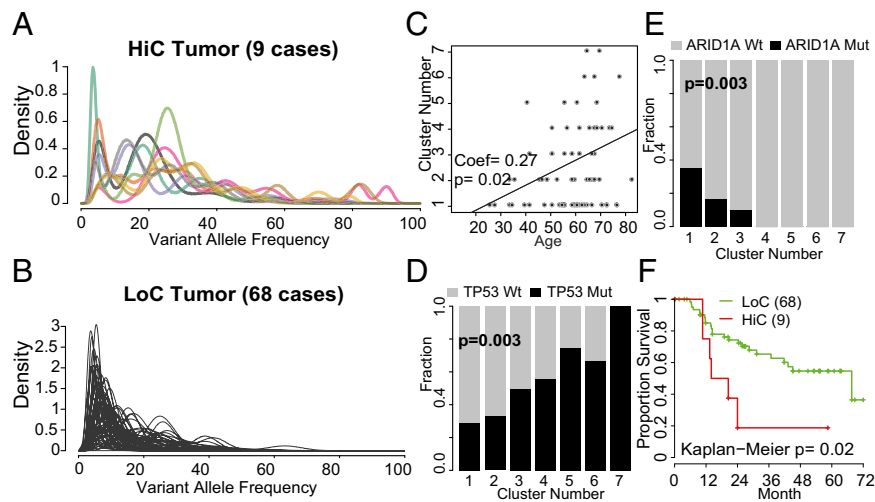
**Identification of High Clonality and Low Clonality Subtypes in GC by Clonality Analysis.** The 78 GC samples in discovery panel and their matching blood cells were subjected to WES. A mean coverage depth of 167X was achieved for GC tumor DNA, and 170X for matched normal blood cells. Eighty-five percent of exons were

covered at 20X or greater depth for both tumor and normal samples (Fig. S1 and Dataset S2). We applied MuTect (13) and VarScan2.2.5 (14) to whole-exome data to detect somatic point and indel mutations. A total of 13,866 mutations were detected in this cohort, including 3,421 synonymous, 8,558 missense, 576 nonsense, 241 splice site, 967 frameshift, and 103 nonframeshift indels. The overall mutation frequency was 112.5 mutations (range of 1–1,486 mutations) with a median of 82.5 for nonsilent mutations per tumor (Dataset S3).

To gain insight into GC heterogeneity, we performed SNV clonality analysis using a recently reported algorithm SciClone (15), which estimates the clusters of clones in each sample. This analysis identified cluster numbers ranging from 1 to 8 in our cohort with a bimodal distribution dipping at a cluster number of 4. We thus define those GC cases with cluster numbers larger than 4 as high clonal or HiC (nine cases) (Fig. 1A and Dataset S4) and cases with cluster numbers equal to or fewer than 4 as low clonal or LoC (68 cases) (Fig. 1B and Dataset S4). Association analyses with clinical parameters revealed that the HiC subtype was associated with older disease onset age (Fig. 1C) (Pearson correlation coefficient = 0.27,  $P = 0.02$ ). There were a total of 11,768 protein-coding subclonal SNVs (92.3% of all detected mutations, with a median of 106 mutations per sample) (Dataset S3). Gene-level mutation analysis revealed that the HiC subtype had a significantly higher frequency of *TP53* mutations (Fig. 1D), whereas the LoC subtype was enriched for *ARID1A* mutations (Fig. 1E), both thought to be the driving events in GC tumorigenesis. Previous studies of leukemia (16) have reported that the higher clonal complexity was associated with poor survival, likely because of chemoresistance. Among our cohort of 78 GC patients, 32 (41.03%) died from GC with a median follow-up of 25.1 months (range from 0.20 to 78.95 months) (Dataset S1). Survival analysis showed that the HiC subtype was significantly associated with shorter survival compared with the LoC subtype ( $P = 0.02$ , log-rank test) (Fig. 1F). There were no significant differences in the treatment regimens between the HiC and LoC subtypes. This association remained significant after adjusting for age, sex, stage, Lauren subtype, *TP53* mutation, and *ARID1A* mutation [adjusted hazard ratio (HR), 4.69; 95% confidence interval (CI), 1.62–13.6;  $P = 0.0043$ ], indicating that the clonal subtype is an independent prognostic factor for GC.

The distinct clinical behavior of the HiC and LoC subtypes suggest they may have fundamental differences in tumorigenesis, as was recently proposed to be manifested by various nucleotide level mutational signatures (17, 18). Detailed mutational signature analysis of these two subtypes showed a significant difference in the percentages of C-to-G transition and C-to-T transversion mutations. HiC GCs had a significantly smaller fraction of C-to-T transversion mutation (45% in HiC vs. 52% in LoC) but greater fraction of C-to-G transition mutation (25% in HiC vs. 9% in LoC,  $P = 0.002$ ). Specifically, the HiC subtype had a significantly higher fraction of C > G mutation in almost all contexts, except for the G in the 5'-adjacent to C (Dataset S4 A–C).

Although *TP53* mutations were commonly found in the HiC group, they were present only in the minor clones (mutation frequency < 15%) in all six positive HiC cases (Dataset S4 D and E). Conceivably, any of the minor clones can become a major clone after a single agent therapy, which has been illustrated in a study in leukemia (16). Thus, in principle, all of the subclones in a tumor should be eliminated for curative therapy. We used a drug-gene interaction database, DGIdb (19), to search for drugable targets for the clones present in the HiC tumors (Dataset S4F). As expected, many gene mutations and the clones harboring them do not have corresponding drugs available. However, for tumors like PGM71 and PGM32, targetable mutations were identified for almost all subclones [e.g., *ALK*, *ABL2*, *SMAD4* (SMAD family member 4), *FANCG*, *NRG1*, *KRAS* (Kirsten rat sarcoma viral oncogene homolog), and others]



**Fig. 1.** Identification of HiC and LoC subtypes by clonality analysis and their clinical outcomes. (A and B) SciClone analysis with cluster number 4 as cut-off identified 9 HiC and 68 LoC cases. Each peak infers one subclone by SciClone algorithm. (C) Association between clonality (y axis) and age at diagnosis (x axis). (D) Fractions of *TP53* mutation carriers (y axis) in different clonality GC (x axis). (E) Fraction of *ARID1A* mutation carriers (y axis) in different clonality GC (x axis). (F) Kaplan–Meier survival curves for HiC and LoC cases.

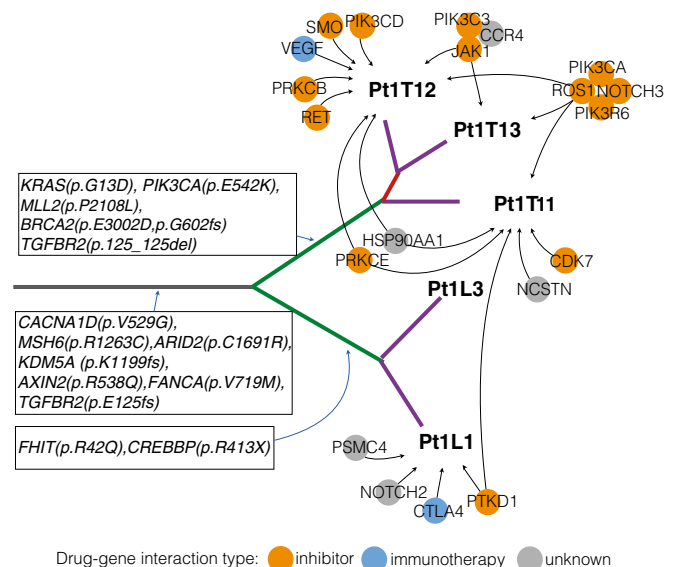
(Dataset S4G). In these cases, combination chemo- and targeted therapies might be much more effective than those using single agents. For other HiC cases the targetable mutations were observed in only two or three clones, underscoring the need for drug development strategies that test multiple combinations (Dataset S4H).

**Clonal Evolution of GC During Metastasis.** To define the type and extent of intratumoral heterogeneity as GC progresses to metastasis, we performed WGS of two cases (Pt1 and Pt2) where, for each case, we obtained the information by dissection and anatomical mapping of three different areas of primary tumor and two lymph node metastases. Pt1 had a total of 4,082 nonsilent mutations, including 2,809 missense, 143 nonsense, 132 splice site, 958 frameshift, and 40 in-frame indels. Pt2 had a lower mutation rate with 287 nonsilent mutations (253 missense, 14 nonsense, 10 splice site, 3 frameshift, and 7 in-frame). Consistent with the higher mutation rate in Pt1, mutations were found in several DNA-mismatch repair and chromatin remodeling genes, including *MSH6*, *TGFBR2*, *KDM5A*, and *MLL4*. In Pt2, the *SETBP1* gene (p.S944N) was mutated, as it is in myeloid malignancies (20), together with amplification of *AKAP9* and *CDK6* (cyclin-dependent kinase 6). Intriguingly, mutations of the gene encoding the voltage-dependent calcium channel protein, *CACNA1D* p.V529G, was found in both Pt1 and Pt2. Copy number alteration analysis showed that a large proportion of the genome was altered in both Pt1 and Pt2, accounting for 516.6 Mb and 706.8 Mb, respectively. Pt2 exhibited consistent genomic amplification of 7q21.2–7q21.3, and Pt1 had amplifications across chr7 and chr8. Pt1 and Pt2 had a comparable number of structural variations with 3,812 and 3,390 breakpoints, respectively.

Because clonal evolution is closely linked to cancer metastasis, we applied phylogenetic tree analysis (*SI Methods*) to the five samples for each GC case and observed considerable divergence among the three primary tumor samples in each (Fig. 2 and Fig. S2). Interestingly, only one part of the primary tumor was closely related to the two highly similar lymph-node metastases, suggesting they were derived from a common clonal ancestor. The multiple clones uncovered within primary tumor and lymph-node metastases suggests the necessity of combined therapy that targets different genetic changes found within the patient's combined tumor burden. We therefore used DGIdb (19) to search for potentially drugable targets in different parts of the same

tumor and their resultant metastases. Results of the drug–sample interaction analysis suggest that Pt1 might benefit from combined targeting of PI3K, CDK7, and Notch (as well as immunotherapy) (Fig. 2), whereas combined AURKC and CDK6 inhibitors might benefit Pt2 (Fig. S2).

**Validated *NRGs-ERBBs* Mutations in GC.** Among the total of 13,866 mutations, we also sought to identify clinically actionable mutations in GC. We analyzed the WES data of the 78 GCs with statistical approaches including MutSigCV (*Methods*) and identified 16 significantly (*q* value < 0.2) mutated genes. Of



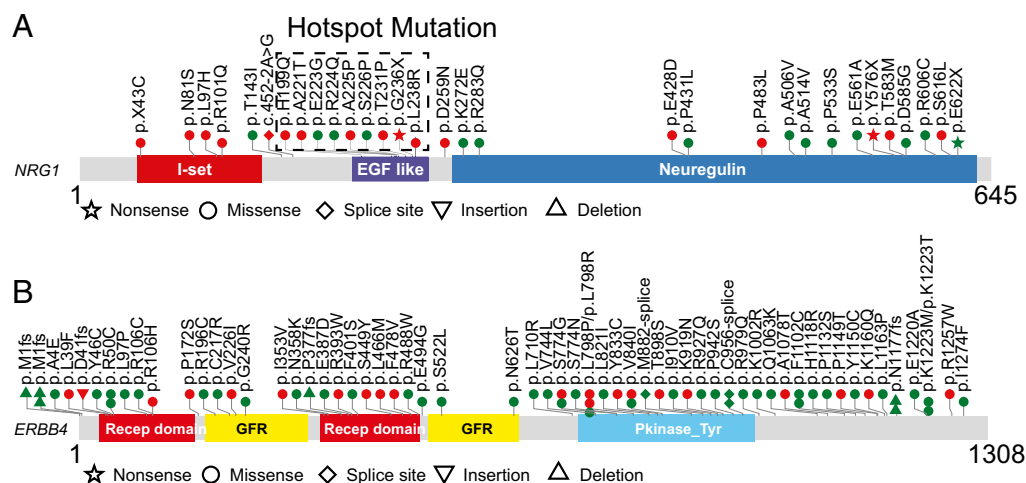
**Fig. 2.** Clonal evolution in GC metastasis revealed by phylogenetic tree. WGS data from five samplings (three primary and two lymph node metastases) of one GC case were used in the similarity analysis of somatic substitutions. Branch and trunk lengths are proportional to the number of nonsynonymous mutations acquired on the corresponding branch or trunk. Driver mutations and copy number events were acquired by the indicated genes in the branches the arrows indicate. The potential drugs which target each tumor portion are shown as nodes with arrows pointing to the samples.

these, 13 were mutated in 5% or more of the tumors. The analysis identified previously reported gene mutations in GC: *TP53*, *ARID1A*, *CDH1* [cadherin 1, type 1, E-cadherin (epithelial)], *APC* (adenomatous polyposis coli), *RHOA*, *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha), *SMAD4*, *MYC* (v-myc avian myelocytomatosis viral oncogene homolog), and *KRAS*. Further pathway analysis of recurrently mutated genes revealed that mutated genes in our cohort belonged to the TP53 pathway, Wnt pathway, ERBB pathway, and HR pathway, and so forth (Dataset S5). Based on mutation frequency and novelty (Dataset S3), we selected 103 recurrently mutated genes in Wnt (Fig. S3), ERBB (Fig. S4A), and HR (Fig. S5) pathways for targeted sequencing verification in an additional 216 GC samples (Dataset S1) using the Ion-Torrent platform (Methods). Among the 103 genes, 86 (83%) were found to have at least one mutation in the validation cohort (Dataset S3 and S6–S10).

The extensive follow-up data associated with our GC cohort allowed us to evaluate the clinical significance of many of the mutations identified in this and previous studies. We first evaluated if there are some genes specifically associated with different anatomical locations in GC. Using 78 WES samples, we identified 10, 3, and 4 genes, which showed moderate enrichment in antrum [i.e., *CSMD3* (30.7%), *TSHZ3* (15.4%), *PCDH11* (19.2%), *ARHGAP28* (15.4%), *DST* (23.1%), *SORCS1* (19.2%), *MUC17* (15.4%), *PCDH20* (19.2%), *USP9X* (19.2%), *CACNA2D1* (15.4%)], body [i.e., *FLG2* (15.4%), *GRM1* (15.4%), *SYNE2* (15.4%)], and cardia [i.e., *CDH1* (23.1%), *GFRAL* (15.4%), *SBNO1* (15.4%), *STAB2* (15.4%)], respectively ( $P < 0.05$ ). Seven (the underlined ones) of those genes remained significant when we included the extra 216 targeted sequencing cases. Intriguingly, none of these gene mutations were observed to significantly associate with same anatomical location in TCGA data, suggesting different genetic background (Russian vs. Chinese) or even diets might contribute to the mutation distribution. In addition to anatomical location, we also interrogated the association between gene mutation and Lauren subtypes of GC. We found that *CDH1* somatic mutations (21 of 294 cases, 7.1%) were strongly enriched in the diffuse GC (Fig. S6A) and 8 of these 21 somatic *CDH1* mutations were loss-of-function types (e.g., nonsense and frameshift mutations). *RHOA* mutations were also enriched in diffuse GC, consistent with recent reports (11, 12). Interestingly, Wnt pathway mutations showed a moderate enrichment among intestinal GC. Survival analysis showed

that the *CDH1*-mutated GCs had shorter survival, whereas Wnt pathway mutated GCs had longer survival (log-rank  $P = 0.04$ ) (Fig. S6B). This difference was not a result of their association with the two Lauren classifications because we did not observe significantly different survival of these latter two groups in our cohort. Interestingly, *RHOA*-mutated cases were not associated with shorter survivals similar to *CDH1*-mutated GCs, suggesting *RHOA* driven and *CDH1* driven diffuse GCs are different clinically.

Although *ERBB2* has been previously recognized as a therapeutic target for GC because of its gene amplification and mutation (confirmed in our study), our analysis revealed that another member of the ERBB pathway, *ERBB4*, was significantly mutated in both our discovery and validation cohorts. Mutation of *ERBB4* or *HER4* has recently been shown to be an oncogenic driver in melanoma (21), although this has not been previously described for GC. Mutated *ERBB4* activates both *ERBB4* and PI3K-AKT signal transduction and can be inhibited by the dual-tyrosine kinase inhibitor Lapatinib, for which efficacy is presently being evaluated in phase III clinical trials in *ERBB4*-mutated melanoma ([clinicaltrials.gov/ct2/show/NCT01264081](http://clinicaltrials.gov/ct2/show/NCT01264081)). In our discovery cohort, *ERBB4* was mutated in seven samples. In the validation cohort, another 13 samples had *ERBB4* mutations. Among these mutations, *ERBB4* p.R50C, was also observed in melanoma (Fig. 3A). Unlike *ERBB2*, which has no known ligand, *ERBB4* binds to NRGs, particularly *NRG1*, forming homo- and heterodimers, and leading to the subsequent activation of kinase activity. Previous studies have shown that overexpression of *NRG1* leads to the activation of ERBBs and consequent Lapatinib-sensitivity (22). However, no somatic mutation of *NRG1* has been reported previously. In our discovery cohort, *NRG1* was mutated in five samples. In the validation cohort, *NRG1* mutations were observed in another 16 cases. Taking these data together, we find that *NRG1* and *ERBB4* were mutated in 34 GC samples (11.6%), with 3 samples having mutations in both genes and 31 samples having mutations in either *NRG1* or *ERBB4* (Fig. 3). Detailed analysis showed that 7 of 21 (33%) *ERBB4* mutations occurred in the kinase domain and 4 (20%) in the receptor domains, suggesting these mutations might impact receptor ligand interaction and kinase activity. Interrogation of the 289 GC cases in the TCGA database also revealed *ERBB4* mutations present in both the receptor domain (i.e., p.R106H/C) and the kinase domain (i.e., p.V744L and p.774N/G). We also found *NRG1* mutations in the TCGA cohort, including the mutations



**Fig. 3.** Frequent *NRG1* and *ERBB4* gene mutations in GC. (A) Distribution of coding mutations in *NRG1* gene. Red: mutations present in Tianjin cohort; green: mutations present in TCGA cohort. (B) Three-dimensional structure for *NRG1* protein and mutation locations.

(i.e., p.A221T and p.A225P) in the EGF-like domain (predicted as the functional domain to bind ERBB4), as well as in regions adjacent to the EGF-like domain (i.e., p.E223G, p.R224Q, and p.S226P). These findings suggest that the EGF-like domain is a hotspot mutation for NRG1. Mutations were also present in other NRGs and ERBB family proteins ([Dataset S11](#)). Mutual exclusivity analysis was conducted for ERBB and NRG gene members in the combined Tianjin and TCGA cohorts. In 1,000-times permutation, the *ERBBs* and *NRGs*' mutations were mutually exclusive of each other ( $P = 0.02$ ).

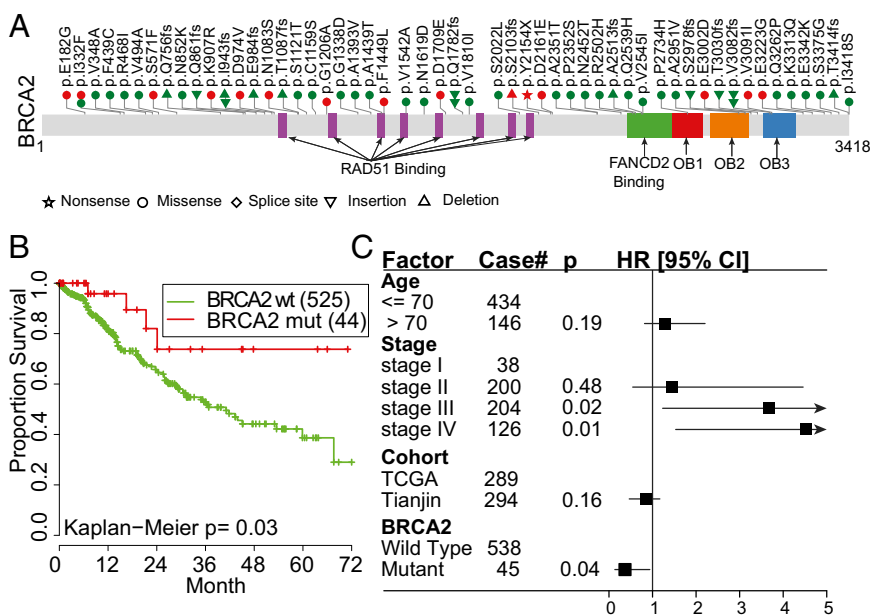
**GCs with *BRCA2* Mutations Show a Trend for Longer Survival.** Adjuvant chemotherapy as the standard of care for GC has limited success and there is no reliable marker to predict response. Among our 78 WES cohort of GC cases, 5 had *BRCA2* mutations (interestingly, there were no *BRCA1* mutations). Correspondingly, in the 216 targeted sequencing cohort cases, another 12 had *BRCA2* mutations (10 somatic and 2 germ line) ([Dataset S3](#) and [S5](#)). A higher frequency (28 of 289 cases) of *BRCA2* mutations was also found in the TCGA Russian GC cohort (Fig. 4A). This finding is consistent with epidemiological studies showing that GC is the third major cancer type, joining breast and ovarian cancer, with *BRCA2* mutations (23). We hypothesized that the *BRCA2* mutation might be clinically important for GC because of accumulating evidence demonstrates that such mutations are associated with improved response to platinum-based chemotherapy and longer survival in ovarian cancer (24). *BRCA2* mutant carrier patients are marginally correlated with longer survival in Tianjin cohort ( $P = 0.07$ , HR = 0.29, 95% CI: 0.07–1.19). A similar nonsignificant trend was also observed in TCGA cohort ( $P = 0.23$ , HR = 0.41, 95% CI: 0.10–1.77). When these two cohorts were pooled to increase the sample size to 569 (and *BRCA2* mutation carriers to 45 cases), mutation carriers were significantly correlated with longer survival (log-rank  $P = 0.03$ ) (Fig. 4B). After adjusting for age, stage, and different population, *BRCA2* mutation remained a significant factor predicting longer survival ( $P = 0.05$ , HR= 0.37, 95% CI: 0.13–0.96) (Fig. 4C).

## Conclusion and Discussion

The most important finding to emerge from this study is the description of intratumoral heterogeneity, which was revealed by clonality analysis. The observation that mutational signatures of two subtypes of GC are distinct from each other indicates there might be fundamental difference in etiology (such as mutagens) of two subtypes, which cause different mutational patterns. The GCs with the highest heterogeneity, which we termed the HiC subtype, have the poorest survival. This result could be because of the escape of specific subclones from the therapeutic effect of single agents. Gene-level analysis further identified *TP53* and *ARID1A* mutation as markers for HiC and LoC, respectively, which may have translational significance in clinic, although this will need to be validated in the future.

Tumor heterogeneity is clearly a hallmark for metastatic tumors, a notion supported by our phylogenetic tree analysis of WGS data from two metastatic GC cases where we were able to analyze multiple samples. Metastasis can be challenging to treat because metastatic cells can be quite different from those of their primary tumors. The drug target analysis of the heterogeneous clones suggested that specific clones have differing drugable gene mutations, yet whether all would need to be targeted for effective therapy remains to be determined. It is plausible that combination therapy that targets driver mutations in all of the key clones within a tumor, potentially combined with immunotherapeutic approaches, might be an effective approach to treat a heterogeneous cancer like GC.

The present study of Chinese GC identified a number of commonly mutated genes that have been reported in other populations (Hong Kong and Russian) including *TP53*, *ARID1A*, *CDH1*, *APC*, *RHOA*, *PIK3CA*, *SMAD4*, *MYC*, and *KRAS*. Our study also revealed some mutation events that were not reported previously. Notably, the *NGR1* and *ERBB4* genes were commonly mutated in our cohort. Whether this is more specific to the Chinese population compared with others remains to be replicated. However, this finding is of potential significance because inhibitors for ERBB signaling are available and in clinical trials. Our study also revealed *BRCA2* mutation as a predictor of longer survival, similar to that



**Fig. 4. *BRCA2* mutation carriers have longer survival.** (A) Distribution of coding mutations in *BRCA2* gene. Red: mutations in Tianjin cohort; green: mutations in TCGA cohort. The number of total mutations is more than the number of patients is because some patient has multiple *BRCA2* mutations. (B) Kaplan-Meier survival curves for *BRCA2* wild-type and *BRCA2* mutated GCs. (C) Forest plot of multivariable Cox-regression model including clinical and *BRCA2* mutations.

reported in ovarian cancer (24). It has been proposed that the reason *BRCA2* mutations are associated with improved survival is that *BRCA2* mutated tumor cells have a diminished capacity to repair damaged DNA caused by chemotherapy. Interestingly, platinum-based chemotherapy is frequently a component of front-line therapy for GC, similar to ovarian cancer. One notable distinction of *BRCA2* mutation profiles in GC is that most are missense mutations (Fig. 4A), whereas mutations in triple-negative breast cancer and ovarian cancer are nonsense or frameshift. This finding suggests that most of the *BRCA2* mutations found in GC are deleterious mutations, although this has to be functionally validated in the future. Nevertheless, our study identified *BRCA2* mutations as a prognostic marker that could be clinically useful for GC management and clinical trial design.

## Methods

**Illumina-Based Whole-Exome Sequencing.** The whole-exome capture procedure was performed with Agilent's SureSelect Human All Exon Kit

protocol. Resulting DNA libraries with an insert size of 200 bp on average were sequenced using the 90-bp paired-end technology on Illumina HiSeq. 2000. Real-time image analysis and base calling were performed by HiSeq Control Software v1.1.37 and Real Time Analysis v1.7.45 using standard parameters, respectively.

**PGM Sequencing.** Next-generation sequencing was performed using the Personal Genome Machine (PGM, LT).

**Statistical Analysis.** Student's *t* test, analysis of variance,  $\chi^2$ , Wilcoxon rank-sum test, Fisher's exact test, Kaplan–Meier estimate, and Mantel–Cox survival analyses were performed using R 2.10.0. Significance was defined as  $P < 0.05$ . Benjamini–Hochberg multiple testing correction was used to estimate the false-discovery rate when multiple testing correction was applied.

**ACKNOWLEDGMENTS.** This research was funded by the grants from The National Foundation for Cancer Research and the Asian Fund for Cancer Research.

1. Ferlay J, et al. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 127(12):2893–2917.
2. Uemura N, et al. (2001) *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* 345(11):784–789.
3. Murphy G, Pfeiffer R, Camargo MC, Rabkin CS (2009) Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology* 137(3):824–833.
4. Shah MA, Ajani JA (2010) Gastric cancer—An enigmatic and heterogeneous disease. *JAMA* 303(17):1753–1754.
5. Catalano V, et al. (2009) Gastric cancer. *Crit Rev Oncol Hematol* 71(2):127–164.
6. Chen W, et al. (2013) The incidences and mortalities of major cancers in China, 2009. *Chin J Cancer* 32(3):106–112.
7. Kim B, et al. (2003) Expression profiling and subtype-specific expression of stomach cancer. *Cancer Res* 63(23):8248–8255.
8. Song F, et al. (2014) Integrated microRNA network analyses identify a poor-prognosis subtype of gastric cancer characterized by the miR-200 family. *Clin Cancer Res* 20(4):878–889.
9. Wang K, et al. (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 43(12):1219–1223.
10. Bass AJ, et al.; Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513(7517):202–209.
11. Wang K, et al. (2014) Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* 46(6):573–582.
12. Kakiuchi M, et al. (2014) Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* 46(6):583–587.
13. Cibulskis K, et al. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31(3):213–219.
14. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.
15. Miller CA, et al. (2014) SciClone: Inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLOS Comput Biol* 10(8):e1003665.
16. Landau DA, et al. (2013) Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152(4):714–726.
17. Alexandrov LB, et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013) Signatures of mutational processes in human cancer. *Nature* 500(7463):415–421.
18. Lawrence MS, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.
19. Griffith M, et al. (2013) DGIdb: Mining the druggable genome. *Nat Methods* 10(12):1209–1210.
20. Makishima H, et al. (2013) Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet* 45(8):942–946.
21. Prickett TD, et al. (2009) Analysis of the tyrosine kinome in melanoma reveals recurrent mutations in ERBB4. *Nat Genet* 41(10):1127–1132.
22. Wilson TR, Lee DY, Berry L, Shames DS, Settleman J (2011) Neuregulin-1-mediated autocrine signaling underlies sensitivity to HER2 kinase inhibitors in a subset of human cancers. *Cancer Cell* 20(2):158–172.
23. Friedenson B (2005) BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. *MedGenMed* 7(2):60.
24. Yang D, et al. (2011) Association of BRCA1 and BRCA2 mutations with survival, chemotherapy sensitivity, and gene mutator phenotype in patients with ovarian cancer. *JAMA* 306(14):1557–1565.