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

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

^aDepartment of Epidemiology and Biostatistics, ^fDepartment of Pathology, and ^hDepartment 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; ^bDepartment of Pathology, University of Texas MD Anderson Cancer Center Informatics Center, Houston, TX 77030; ^cCenter of Pharmacogenetics, Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15261; ^dBeijing Genomics Institute-Shenzhen, Shenzhen 518083, Guangdong, People's Republic of China; ^eDepartment 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; ^gDepartment of Oncology Discovery, Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322; ⁱDivision of Medical Oncology, Mayo Clinic, Rochester, MN 55905; ^jMedical Sciences, Amgen Inc., Thousand Oaks, CA 91320; ^kLilly China Research and Development Center and ^lTranslational Sciences Oncology, Eli Lilly and Company, Indianapolis, IN 46285; ^mTranslational Oncology & Personalized Medicine, AbbVie Inc., North Chicago, IL 60064; ⁿExploratory Clinical & Translational Research, Bristol-Myers Squibb India, Mumbai 400013, India; ^oApplied Genomics and Bioinformatics, Bristol-Myers Squibb, Princeton, NJ 08540; ^pLudwig Institute for Cancer Research, University of California at San Diego, La Jolla, CA 92093; ^qHuman Cancer Genetics Program, Ohio State University, Columbus, OH 43210; ^rThe National Foundation for Cancer Research, Bethesda, MD 20814; ^sAsian Fund for Cancer Research, Hong Kong, People's Republic of China; and ^tDepartment 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

Gastric 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 (accession no. EGAS00001001056).

¹K.C., D.Y., and X. Li contributed equally to this work.

²To whom correspondence may be addressed. Email: wcavenee@ucsd.edu, xishanhao@sina.com, or wzhang@mdanderson.org.

This article contains supporting information online at 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]