

# Whole-exome sequencing of alpha-fetoprotein producing gastric carcinoma reveals genomic profile and therapeutic targets

Jun Lu<sup>1,7</sup>, Yongfeng Ding<sup>2,7</sup>, Yanyan Chen<sup>1,7</sup>, Junjie Jiang<sup>1</sup>, Yiran Chen<sup>1</sup>, Yingying Huang<sup>1</sup>, Mengjie Wu<sup>1</sup>, Chengzhi Li<sup>3</sup>, Mei Kong<sup>3</sup>, Wenyi Zhao<sup>4</sup>, Haohao Wang<sup>1</sup>, Jing Zhang<sup>1</sup>, Zhongqi Li<sup>1</sup>, Yimin Lu<sup>1</sup>, Xiongfei Yu<sup>1</sup>, Ketao Jin<sup>1</sup>, Donghui Zhou<sup>1</sup>, Tianhua Zhou<sup>5</sup>, Fei Teng<sup>6</sup>, Haibin Zhang<sup>1</sup>, Zhan Zhou<sup>4</sup> <sup>✉</sup>, Haiyong Wang<sup>1</sup> <sup>✉</sup> & Lisong Teng<sup>1</sup> <sup>✉</sup>

Alpha-fetoprotein producing gastric carcinoma (AFPGC) is a rare and aggressive subtype of gastric cancer. However, little is known about the genomic features of this disease. We perform whole-exome sequencing analysis of AFPGC, and identify 34 significantly mutated genes. Somatic copy number alterations analysis reveals several significant focal amplifications (e.g. 19q12, 17q12) and focal deletions (e.g. 1p36.11, 9p21.3), and some of these negatively affect the patient prognosis. Comparative analyses reveal that AFPGC has distinct genomic features from gastric cancer of The Cancer Genome Atlas as well as four molecular subtypes. Several frequently altered genes with potential as therapeutic targets are identified in AFPGC. Further analysis reveals that AFPGC with amplification of *CCNE1* at 19q12 and/or *ERBB2* at 17q12 show poorer survival and more aggressive. Subsequently, based on our established patient-derived xenograft models for AFPGC, translational research is performed and the therapeutic value of targeting *CCNE1* and *ERBB2* is validated. In this work, we provide an understanding of genomic characteristics of AFPGC and propose a platform to explore and validate the genome-guided personalized treatment for this disease.

<sup>1</sup>Department of Surgical Oncology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>2</sup>Department of Medical Oncology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>3</sup>Department of Pathology, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>4</sup>Innovation Institute for Artificial Intelligence in Medicine and Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences and Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Zhejiang University, Hangzhou, China. <sup>5</sup>Institute of Gastroenterology, Cancer center, Zhejiang University, Hangzhou, China. <sup>6</sup>Hangzhou Oncocare Co. Ltd, Hangzhou, China. <sup>7</sup>These authors contributed equally: Jun Lu, Yongfeng Ding, Yanyan Chen. ✉email: [zhanzhou@zju.edu.cn](mailto:zhanzhou@zju.edu.cn); [lanceter1@zju.edu.cn](mailto:lanceter1@zju.edu.cn); [lsteng@zju.edu.cn](mailto:lsteng@zju.edu.cn)

**α**-Fetoprotein (AFP)-producing gastric carcinoma (AFPGC) is a rare and special subtype of gastric cancer (GC) associated with high liver metastasis rate and extremely poor prognosis<sup>1,2</sup>. According to the World Health Organization (WHO) classification of digestive system tumors (5th edition, 2019)<sup>3</sup>, AFPGC is characterized as a rare type of GC with elevated serum AFP levels and positive immunohistochemistry (IHC) staining for AFP, accounting for 1.3–5.4% of GC<sup>4–7</sup>.

To date, omics analyses of GC are performed to identify molecular characteristics of GC. The Cancer Genome Atlas (TCGA) investigators have published comprehensive genomic and transcriptomic analyses of GC and identified four subtypes: Epstein–Barr virus-infected (EBV), microsatellite instability (MSI), genomically stable (GS), and chromosomally unstable (chromosomal instability, CIN) tumor types<sup>8</sup>. Subsequently, the investigators of the Asian Cancer Research Group (ACRG) categorized GC into another four subtypes with differential prognosis<sup>9</sup>. Recently, some studies have focused on characterizing the molecular signatures of several special subtypes of GC, such as diffuse-type GC<sup>10</sup>, early-onset GC<sup>11</sup>, and EBV-positive GC<sup>12</sup>. These clinicopathological subtypes and their associated molecular characterization provided a foundation for better patient stratification and choice for targeted therapy. AFPGC is associated with more aggressive behavior and poorer prognosis as compared to common GC; however, there is no specialized treatment till date and lacked large-scale genomic studies. Thus, it is urgent to elucidate the molecular features and facilitate the development of specialized therapies for AFPGC.

Here we present the molecular landscape of AFPGC by performing whole-exome sequencing (WES), fluorescence in situ hybridization (FISH), and IHC in a cohort of 105 AFPGC patients till now. These results suggest aggressive behavior and distinct genomic features of AFPGC when compared to common GC. We also identify several frequently altered genes that are potentially targetable and finally evaluate these in the corresponding patient-derived xenograft (PDX) models. Our study elucidates the properties of AFPGC and provide a rationale for the development of specialized treatment.

## Result

**Patients and tumor samples.** A total of 5261 GC cases were diagnosed in our institution and then a cohort of 105 GC patients who were IHC positive for AFP were enrolled in our study (Supplementary Fig. 1). Of these, 58 paired fresh-frozen and/or formalin-fixed paraffin-embedded (FFPE) tumor tissues and adjacent normal tissues that met our criteria have undergone WES. All 105 FFPE tumor tissues underwent FISH and IHC tests. Among the 105 AFPGC patients, 79 (75.2%) were males and the median age at diagnosis was 64 years (range: 30–83 years) (Supplementary Data 1 and Supplementary Table 1). The majority (68.6%) of the patients were diagnosed at an advanced stage (Stage I/II, 33 patients and stage III/IV, 72 patients). Liver metastasis was found in 42 (40%) patients in our study. Compared to stage-matched non-AFPGC patients in our institute, AFPGC patients were found to have poorer prognoses (hazard ratio (HR) = 1.47,  $P = 0.015$ ) (Supplementary Fig. 2a). The association between serum AFP level and prognosis of AFPGC was further investigated, and 270 ng/mL was chosen as a cutoff value according to the receiver operating characteristic curve. The results revealed that patients with high serum AFP level (>270 ng/mL) have poorer overall survival (OS) ( $P = 0.017$ ) (Supplementary Fig. 2b, c).

**Somatic mutations and signatures of AFPGC.** WES analysis was performed on the genomic DNA of 58 GC tumors and matched

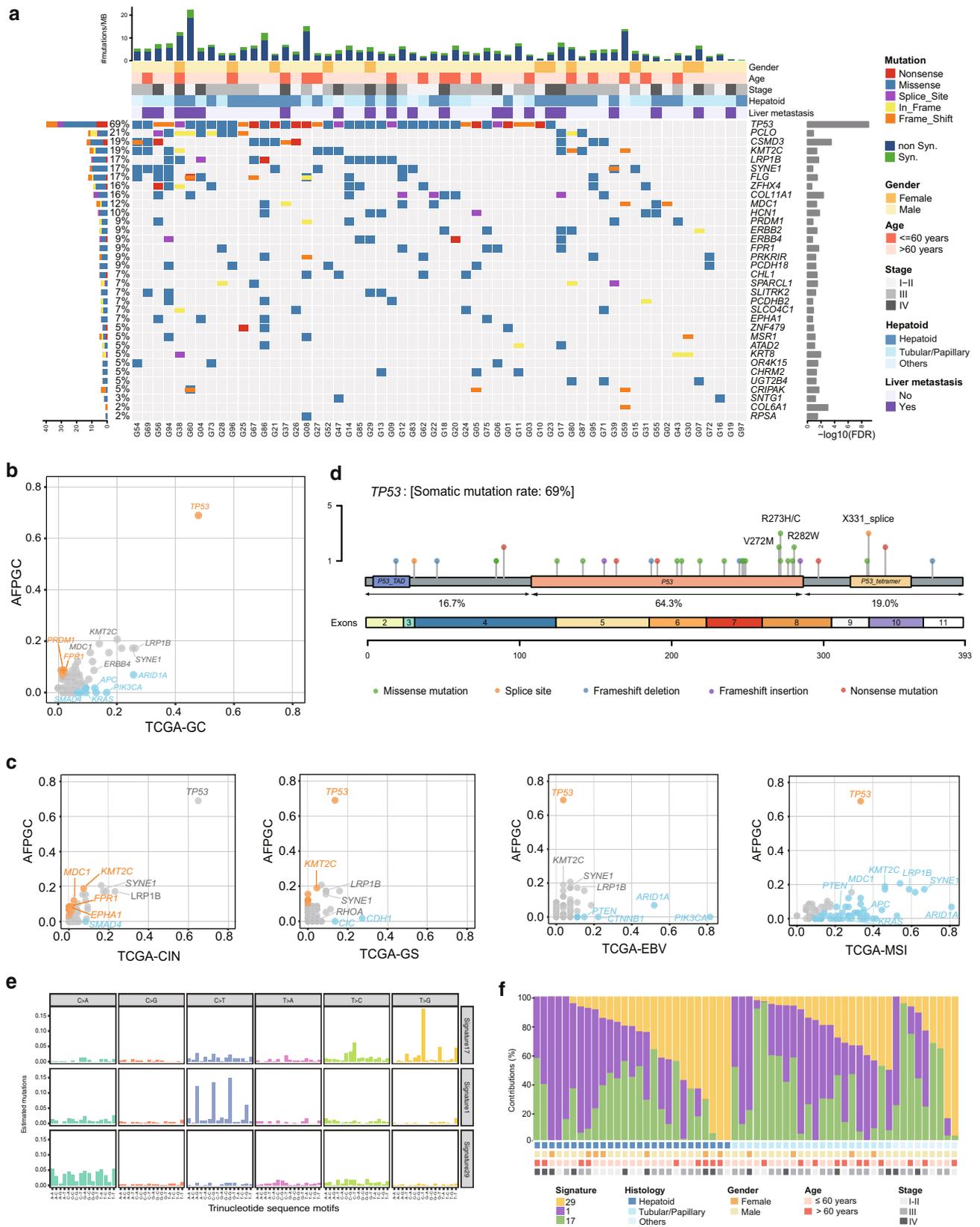
normal tissues at a mean coverage of 161- and 105-fold, respectively (Supplementary Data 2). A total of 1041 somatic single-nucleotide variants (SNVs) and 300 somatic indels were identified, including 1179 synonymous and 161 nonsynonymous mutations that are corresponding to 4.08 nonsynonymous mutations per megabase of the targeted DNA (Supplementary Data 3). Thirty-four significantly mutated genes were identified in AFPGC using MuSic analysis<sup>13</sup> (Fig. 1a and Supplementary Table 2). Among these, the most frequently altered genes in AFPGC were *TP53* (69%), *PCLO* (21%), *CSMD3* (19%), and *KMT2C* (19%). Other frequently mutated genes (>9%) were also detected, such as *LRP1B*, *SYNE1*, *FPRI*, *PRDM1*, and *PRKRIR*. Notably, frequently mutated driver genes (such as *PIK3CA*, *KRAS*, *APC*) in TCGA-GC were rarely mutated in AFPGC (Fig. 1b and Supplementary Data 4). The frequency of *TP53* mutation in AFPGC was significantly higher than that in TCGA-GC ( $P < 0.01$ ). Furthermore, when comparing to TCGA subtypes, *TP53* mutation rate of AFPGC was significantly higher than all subtypes other than TCGA-CIN (Fig. 1c and Supplementary Data 4). Detailed analysis showed that 64.3% of *TP53* mutations detected in AFPGC occurred in the DNA-binding domain, including recurrent missense mutations R273H/C, R272M, and R282W (Fig. 1d). In addition, c.994-1 G > A (X331\_splice), a splice site mutation in the oligomerization domain, was present in three cases of AFPGC but absent in TCGA-CIN or TCGA-GC (Supplementary Fig. 3). On the contrary, *TP53* R175H/G, the most frequent mutation in TCGA-CIN, was not observed in AFPGC. Besides, the mutation rates of other significantly mutated genes, such as *KMT2C*, *MDC1*, *FPRI*, *EPHA1*, and *SMAD4*, were different between AFPGC and TCGA-CIN (Fig. 1c and Supplementary Data 4).

The predominant somatic mutation types were C:G>T:A transitions and C:G>A:T transversions (Supplementary Fig. 4). A total of three independent mutation signatures were screened out (Fig. 1e and Supplementary Data 5). Among these, Signature 1 showed association with a clock-like mutational process and was correlated with age<sup>14</sup>. Signature 29 was found in the cancer samples obtained from individuals with a tobacco chewing habit. Signature 17 was the hallmark signature of esophageal cancer and GC<sup>15</sup>. Importantly, the recurrent pattern of Signature 6 (associated with defective DNA mismatch repair) was found to be absent, which was consistent with the result that no MSI case was found in AFPGC (Supplementary Data 6).

Moreover, the contributions of these mutation signatures in AFPGC across various clinicopathological characteristics were shown in Fig. 1f and no significant relevance was observed (Supplementary Fig. 5).

**Somatic copy number alterations of AFPGC.** In total, 58 paired AFPGC tumor and normal samples were analyzed for somatic copy number alterations (SCNAs) using GISTIC 2.0. SCNAs were recurrent in AFPGC and also included amplifications containing *CCNE1* (19q12), *ERBB2* (17q12), *CCND1* (11q13.3), *MUC4* (3q29), *MCL1* (1q21.3), and *MYC* (8q24.21), and deletion of *ARID1A* (1p36.11), *CDKN2A* (9p21.3), *CREBBP* (16p13.3), and *SMAD4* (18q21.2) (Fig. 2a and Supplementary Data 7). Furthermore, the impact of SCNA on survival outcomes was explored in AFPGC using Kaplan–Meier survival analysis. The results indicated that patients with amplification in 19q12 or 17q12 might have a significantly worse OS rate than those without amplification (19q12,  $P = 0.008$ ; 17q12,  $P = 0.038$ , Supplementary Fig. 6). In addition, patients with 8q24.21 amplification tended to show poorer prognosis ( $P = 0.112$ ).

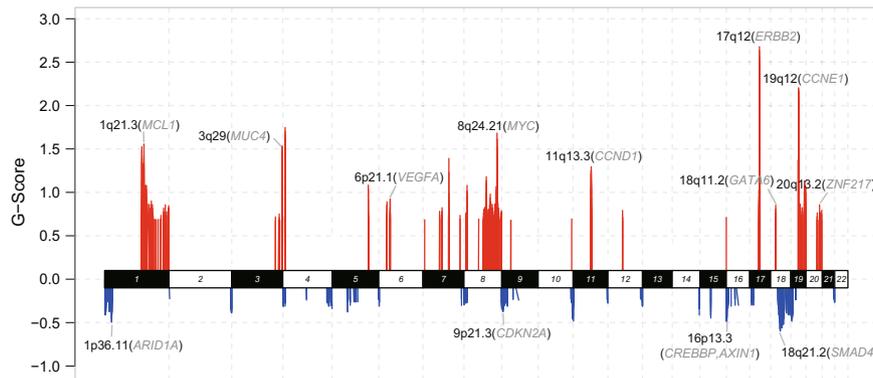
The comparison of significant SCNAs between AFPGC and TCGA-GC was further performed (Supplementary Fig. 7). A total



of 92 significant SCNA regions (42 amplifications and 50 deletions) were identified in AFPGC, of which 58 regions (22 amplifications and 36 deletions) were not significant in TCGA-GC, suggesting a relative specificity for AFPGC (Supplementary Fig. 8a, b). Considering that AFPGC harbored relatively higher number of TP53 mutation rate and SCNAs, we subsequently

compared the SCNAs of AFPGC with TCGA-CIN (a known subtype with recurrent SCNAs in TCGA-GC), finding that 57% (24/42) significant amplifications and 74% (37/50) deletions only existed in AFPGC (Supplementary Fig. 8c–f), such as amplifications in 7q11.21, 6p21.33, and 19q13.43, and deletions in 9q34.3 and 1p36.32 (Supplementary Fig. 7). These regions contain

**Fig. 1 The landscape of somatic mutations and mutational signatures of AFPGC.** **a** Somatic mutations of 58 pairs of AFPGC samples. The matrix in the middle panel shows somatic mutations by tumor sample (column) and by gene (row). The top histogram shows the frequency of nonsynonymous mutation and synonymous mutation. The top tracks show histopathological characteristics of tumor samples. The left histogram shows the number of alterations accumulated on 34 significantly mutated genes identified by MuSic analysis. The right histogram shows negative log transformation of *P*-value. **b, c** Gene mutation rates of AFPGC in comparison with TCGA-GC (**b**) or four subtypes of TCGA-GC (**c**). Orange dots, genes with significantly higher mutation rate in AFPGC; blue dots, genes with significantly lower mutation rate in AFPGC. **d** Distribution of nonsynonymous somatic TP53 mutations in 58 AFPGC. **e** Ninety-six substitutions derived from WES data of 58 pairs of AFPGC samples. The horizontal axis represents mutation patterns of 96 substitutions with different colors. The vertical axis depicts the estimated mutations attributed to a specific mutation type. **f** The distribution of mutation signatures in AFPGC across various clinicopathological characteristics. AFPGC,  $\alpha$ -fetoprotein-producing gastric carcinoma; CIN, chromosomal instability; EBV, Epstein-Barr virus; FDR, false discovery rate; GS, genomically stable; Hepatoid, hepatoid differentiation; MSI, microsatellite instability; Non syn, nonsynonymous mutation; Syn, synonymous mutation; TCGA, The Cancer Genome Atlas; Tubular/Papillary, tubular or papillary differentiation. Source data are provided as a Source Data file.



**Fig. 2 The landscape of somatic copy number alterations of AFPGC.** Somatic copy number alterations identified by GISTIC 2.0 analysis. The genome is oriented horizontally and GISTIC scores at each locus are plotted from the bottom to top. Red represents focal amplification and blue represents deletion in the genomic regions. Representative amplification/deletion and corresponding genes are labeled as the SCNA region (representative gene). G-score, GISTIC scores. Source data are provided as a Source Data file.

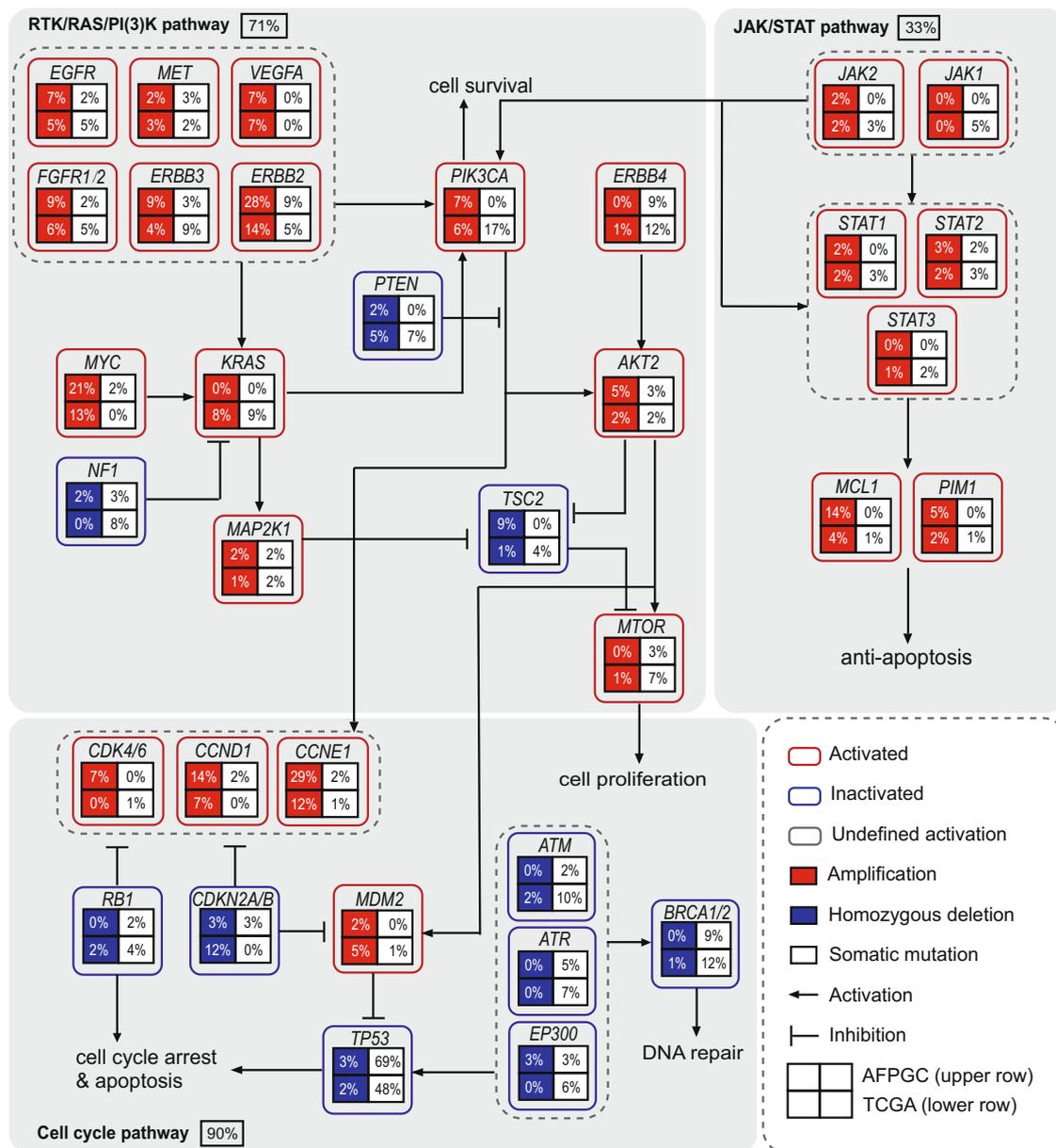
important cancer-related genes including *TRIM28*, *SEMA3C*, *NOTCH1*, and *FAT1*.

**Pathway analysis.** Integrated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of somatic mutations and SCNA data revealed core signaling pathway alterations in AFPGC, including RTK/RAS/PI(3)K, p53/cell cycle, and JAK/STAT (Fig. 3 and Supplementary Data 8). The RTK/RAS/PI(3)K signaling pathway was primarily altered by amplification of *ERBB2*, *ERBB3*, and *MYC*. P53/cell cycle signaling pathway was altered mainly owing to frequent mutations of *TP53* and amplification of *CCNE1/CCND1*, which encodes cyclins. Downstream signaling cascade genes in JAK/STAT signaling pathway, with *MCL1* as the frequently amplified gene, leads to an anti-apoptotic process. Compared to GC in the TCGA cohort (TCGA-GC), both AFPGC and TCGA-GC showed enrichment in the alterations of RTK/RAS/PI(3)K and p53/cell cycle signaling pathways, but in different patterns. In the RTK/RAS/PI(3)K signaling pathway, AFPGC showed recurrent *ERBB2* amplifications and mutations. The key components of the RAS/PI(3)K signaling pathway—*KRAS*, *PIK3CA*, and *PTEN*—showed frequent mutations in TCGA-GC, but were absent in AFPGC. In the p53/cell cycle pathway, cell cycle mediators (such as *CCNE1*, *CCND1*, and *CDK4/6*) were frequently activated in AFPGC, whereas to a significantly lesser degree in TCGA-GC. More *TP53* mutations were harbored in AFPGC, but there were fewer mutations of upstream mediators (*ATM/ATR/EP300*) and DNA repair regulators (*BRCA1/2*) as compared to TCGA-GC.

**Comparative analysis and survival analysis of potentially targetable genes.** According to the genomic alterations in AFPGC,

there were a number of alterations in known cancer-related or potentially targetable genes, including *ERBB2* (31%), *CCNE1* (29%), *MYC* (22%), *CCND1* (16%), *MCL1* (14%), *FLT1* (12%), *ERBB3* (10%), *AURKA* (10%), *AXL* (9%), *BCL6* (9%), *BRCA2* (9%), *EGFR* (9%), *ERBB4* (9%), *FGFR2* (9%), and so on, and these were comparable to a certain extent with that of the TCGA database (Fig. 4a). Representative clinical trials or preclinical studies of potential therapies targeting these genes were listed in Supplementary Table 3. Of these, *ERBB2* and *CCNE1* were found to have the highest alteration rates (31% and 29%) in AFPGC and were significantly higher than those (19% and 13%) present in the TCGA cohort (Supplementary Table 4, both  $P < 0.05$ ). Taken collectively, genomic alterations of potentially targetable genes (top 5) occurred in ~74.1% of the entire AFPGC cohort (Fig. 4b).

Noting that *ERBB2* and *CCNE1* were two well-known oncogenes in 17q12 and 19q12, respectively, the status of *ERBB2* and *CCNE1* was examined by FISH or IHC in a larger clinical cohort ( $N = 105$ , Fig. 4c). The *ERBB2*-positive rate in AFPGC was significantly higher than that in TCGA-GC and subtypes (Fig. 4d and Supplementary Table 5). A similar trend was also observed with regard to *CCNE1*. Furthermore, the prognostic value of *ERBB2* and *CCNE1* in 105 AFPGC patients was explored. The results revealed that the patients with a positive status of *ERBB2* or *CCNE1* had worse survival outcomes when compared to those with a negative status in AFPGC (HR = 2.07,  $P = 0.009$ , Fig. 4e; HR = 2.64,  $P < 0.001$ , Fig. 4g). However, such phenomenon was not found in the TCGA-CIN cohort, which harbors the most frequent *ERBB2* and *CCNE1* amplification in TCGA subtypes (Fig. 4f, h). Moreover, the associations of *ERBB2* and *CCNE1* with clinicopathological features in AFPGC were evaluated (Supplementary Table 6). The positive status of *ERBB2* showed an association with advanced tumor-node-metastasis (TNM) stage

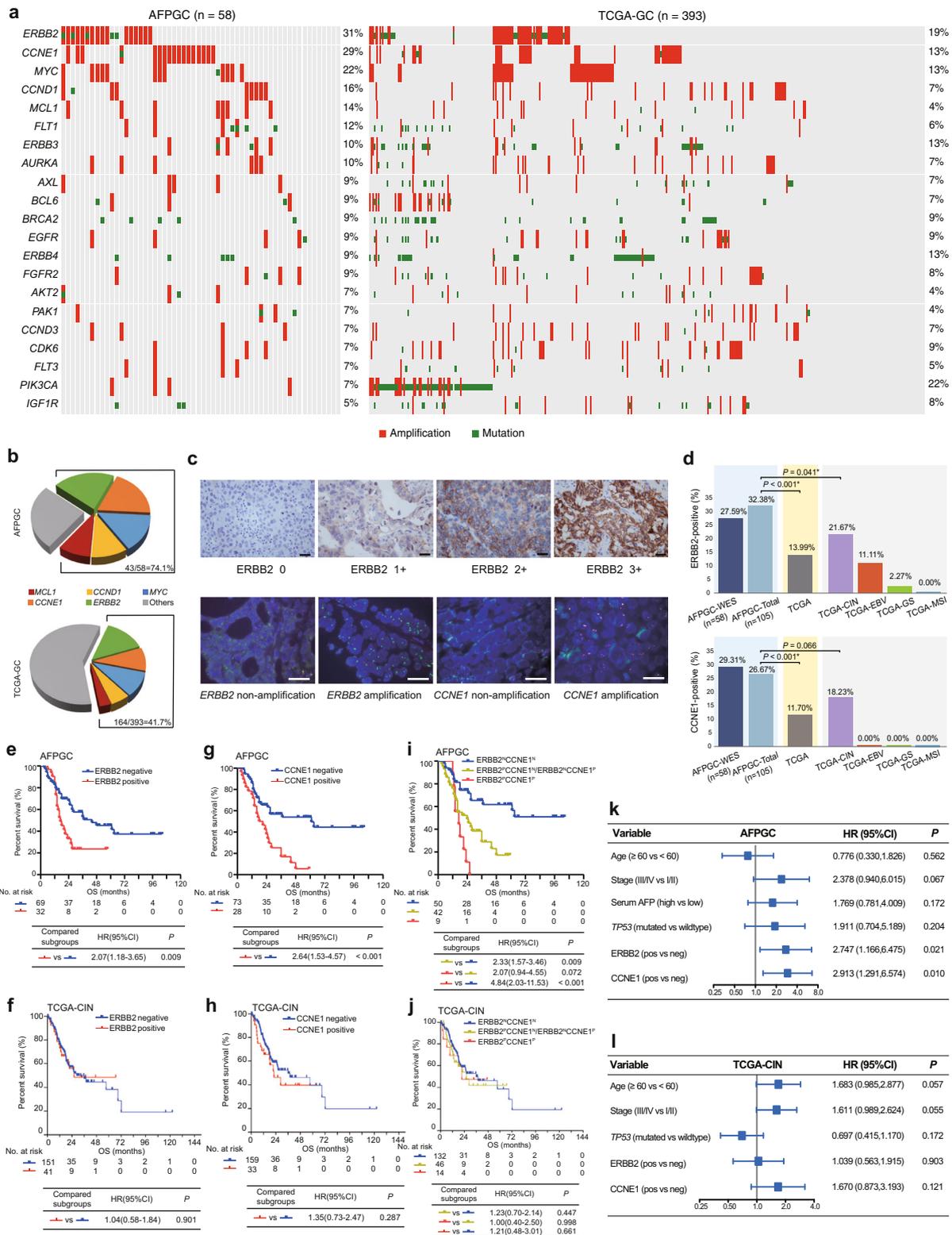


**Fig. 3 Comparison of genetic alterations of signaling pathways in AFPGC and TCGA-GC.** Somatic mutations and SCNAs in key signaling pathways are shown for AFPGC ( $n = 58$ ) and TCGA-GC ( $n = 393$ ) cohort. Three major signaling pathways, namely RTK/RAS/PI(3)K pathway, cell cycle pathway, and JAK/STAT pathway, were shown. In a square grid, the upper row represents AFPGC and the lower row represents TCGA-GC. Percentages represent the alteration frequencies. SCNAs, somatic copy number alterations. Source data are provided as a Source Data file.

( $P = 0.035$ ), positive lymphovascular invasion ( $P = 0.033$ ), and high liver metastasis ratio ( $P = 0.025$ ). Similarly, patients with a positive status of *CCNE1* demonstrated higher ratio of lymphovascular invasion ( $P = 0.012$ ) and liver metastasis ( $P = 0.003$ ). These findings indicated that amplification in *ERBB2* or *CCNE1* might contribute to tumor progression in AFPGC. To examine the combined effects of *ERBB2* and *CCNE1* on prognosis, the patients ( $n = 105$ ) were divided into three groups according to the status of *ERBB2* and *CCNE1* (Co-positive group: *ERBB2* positive and *CCNE1* positive; Single positive group: *ERBB2* positive or *CCNE1* positive; and Co-negative group: *ERBB2* negative and *CCNE1* negative). Interestingly, the co-positive group demonstrated the worst OS rate and aggressive biological characteristics (Fig. 4i and Supplementary Table 7). Although *ERBB2* and *CCNE1* amplification were also enriched in TCGA-CIN, the combined effects of *ERBB2* and *CCNE1* on prognosis were not observed (Fig. 4j).

Furthermore, multivariate Cox regression analysis demonstrated a positive status of *ERBB2* or *CCNE1* was an independent prognostic factor in AFPGC (after adjusting by *TP53* mutation status) but not in TCGA-CIN (Fig. 4k, l).

**Validation of potential therapeutic targets in PDX models.** A series of passable PDX models of AFPGC were successfully established. The status of AFP, *ERBB2*, and *CCNE1* were confirmed by FISH and/or IHC (Fig. 5a and Supplementary Fig. 9) and Fig. 5b–d showed the representatives of AFPGC PDX models with different status of *ERBB2* and *CCNE1*. The clinicopathological features and detailed alterations for each PDX model are summarized in Supplementary Table 8 and Supplementary Fig. 10. *CCNE1*, along with its catalytic subunit Cyclin-dependent kinase 2 (*CDK2*), plays a critical role in the



regulation of G1-S phase transition. Targeting CDK2 has been validated to be an effective anticancer strategy for *CCNE1*-amplified cancer<sup>16,17</sup>. The efficacy of AZD5438 (inhibitor of CDKs, including CDK2) in five AFPGC PDX models with or without *CCNE1* amplification was evaluated. The results showed statistically significant differences in tumor volume between the control group and the AZD5438-treated group in *CCNE1*-amplified AFPGC PDX models, but not in non-*CCNE1*-amplified models (Fig. 5e-h). Next, the antitumor activity of trastuzumab

was evaluated. Consistent with the previously published report, trastuzumab exerted antitumor effects in ERBB2-positive AFPGC PDX models (Fig. 5i, j). Furthermore, the antitumor activity of AZD5438 and trastuzumab was evaluated on ERBB2-positive PDX AFPGC models. According to the previous study, *CCNE1* amplification or overexpression participated in the resistance to trastuzumab treatment in breast cancer<sup>18</sup>. Interestingly, statistically significant differences were also present in the tumor volume between the group treated with AZD5438 combined with

**Fig. 4 Potentially targetable alterations in AFPGC and TCGA-GC, and survival analysis.** **a** Genomic alterations of potentially targetable genes in AFPGC and TCGA-GC. **b** Overall frequency of selected targetable genes in AFPGC and TCGA-GC. **c** Representative images of *ERBB2* and *CCNE1* status. Presented data are a representative image of three independent experiments. Upper panel shows IHC score of *ERBB2* with 0, 1+, 2+, and 3+. Lower panel shows representative FISH image of *ERBB2* and *CCNE1*. Scale bar represents 50  $\mu$ m. **d** The positive rate of *ERBB2* and *CCNE1* in AFPGC and TCGA-GC. **e, f** Associations between *ERBB2* status and OS in AFPGC (**e**) or in TCGA-CIN (**f**). **g, h** Associations between *CCNE1* status and OS in AFPGC (**g**) or in TCGA-CIN (**h**). **i, j** Associations between combined status of *ERBB2* and *CCNE1* and OS in AFPGC (**i**) or in TCGA-CIN (**j**). **k, l** Forest plot of multivariable Cox proportional hazard regression in AFPGC (**k**) or in TCGA-CIN (**l**). The hazard ratios are presented and the horizontal lines indicate the 95% confidence intervals. In AFPGC-WES and TCGA-GC, the positive status of *ERBB2/CCNE1* was defined as amplification by GISTIC 2.0. In AFPGC-total, the definition of the positive status of *ERBB2* and *CCNE1* was shown in Supplementary Methods. Statistical significance was determined using two-sided  $\chi^2$ -test (**d**), log-rank (Mantel-Cox) test (**e–j**), and multivariate Cox regression (**k, l**). AFPGC-WES, the AFPGC cohort with WES data; AFPGC-total, the whole AFPGC cohort in this study; *CCNE1*<sup>N</sup>, *CCNE1* negative; *CCNE1*<sup>P</sup>, *CCNE1* positive; CI, confidence interval; *ERBB2*<sup>N</sup>, *ERBB2* negative; *ERBB2*<sup>P</sup>, *ERBB2* positive; FISH, fluorescence in situ hybridization; HR, hazard ratio; IHC, immunohistochemistry; neg, negative; OS, overall survival; pos, positive. Source data are provided as a Source Data file.

trastuzumab and those treated with AZD5438 or trastuzumab alone in *ERBB2* and *CCNE1* co-amplified PDX models, but the results were not observed in *CCNE1* non-amplified PDX models (Fig. 5i–k).

The treatment efficiency was further evaluated by western blotting analysis in *ERBB2* and *CCNE1* co-positive PDX model after drug administration (Fig. 5l). The results showed that AZD5438 exerted an antitumor effect by inhibiting the expression of CDK2 and phosphorylated retinoblastoma (p-Rb). Moreover, the expressions of p-ERK, p-AKT, CDK2, and p-Rb were greatly decreased in the combined trastuzumab and AZD5438 treatment group as compared to other groups.

## Discussion

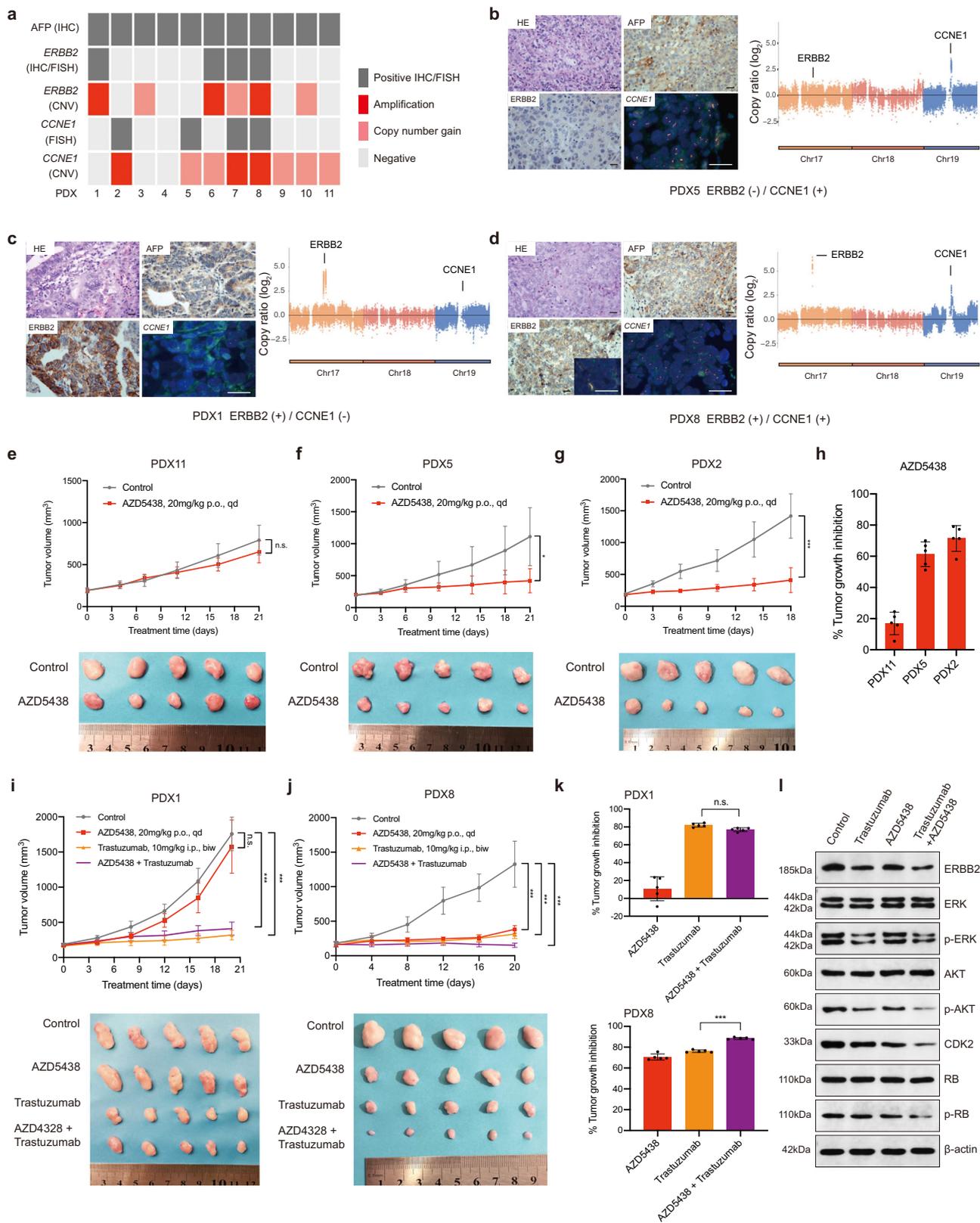
In this study, we provided an understanding of the clinicopathological and molecular features of AFPGC and focused on targetable genomic alterations. We validated that AFPGC had a higher liver metastasis rate and poorer prognosis. Frequent genetic alterations, as well as key signaling pathways that might contribute to the tumorigenesis and development of AFPGC, were revealed. Also, the potentially targetable genes in AFPGC were analyzed and translational research was performed to explore the precise target therapy on PDX models for this disease.

AFPGC was first described by Bourrille et al.<sup>19</sup> in 1971 and was characterized as a special subtype of GC according to the WHO classification of digestive system tumors. Histologically, AFPGC can be identified as hepatoid adenocarcinoma<sup>20</sup> and other differentiated adenocarcinomas, such as tubular/papillary adenocarcinoma, enteroblastic adenocarcinoma, and yolk-sac tumor-like carcinoma<sup>21–25</sup>. More than one of these histological types often coexists in AFPGC.

A study conducted by Hirajima et al.<sup>4</sup> has indicated liver metastasis as a dominant character (which occurred in nearly 40% of AFPGC), as well as an independent prognostic factor in AFPGC<sup>4</sup>. Liu et al.<sup>5,26</sup> have conducted a study on a clinical cohort of 104 AFPGC patients, which showed a high incidence of liver metastasis rate (60%) and dismal prognosis. Similarly, a liver metastasis rate of 40.1% was found in our cohort of 105 AFPGC patients. In addition, the prognosis of AFPGC patients was poorer than those with stage-matched non-AFPGC. However, so far there is no specialized treatment for AFPGC either in the current Japanese Gastric Cancer Association guidelines or National Comprehensive Cancer Network guidelines<sup>27,28</sup>. Till date, very little is known regarding the molecular characteristics of AFPGC that seriously restricted the understanding and management of this disease.

Based on WES results in our study, the significantly mutated genes of AFPGC, such as *TP53*, *CSMD3*, *FPR1*, *PCLO*, *KMT2C*, *SYNE1*, and *LRP1B* were shown. *TP53* acts as a key tumor suppressor that plays a major role in preserving genomic stability<sup>29</sup>.

*TP53* mutations are associated with increased amplification of oncogenes, deep deletions of tumor suppressor genes, and nucleotide-level mutation rates. In our analysis, *TP53* was more frequently mutated in AFPGC when compared to TCGA-GC (most of the cases were negative for AFP<sup>30</sup>). Despite the similarity of the mutation rate, the distribution of *TP53* mutations in AFPGC was different from TCGA-CIN. For instance, three cases in AFPGC harbored X331\_splice mutation, which was absent in TCGA-GC but present in other cancer types including lung, colon, and liver cancers<sup>31</sup>. This mutation was predicted as “disease causing” by MutationTaster<sup>32</sup> and may induce loss of capacity for oligomerization, leading to partial or complete loss of transactivation potential of p53<sup>33</sup>. Notably, as the most frequent hotspot in TCGA-GC (or TCGA-CIN), *TP53* R175 mutation is absent in AFPGC. This mutation was reported to reduce the DNA-binding capacity of p53 (loss-of-function) and was oncogenic in multiple cancers including GC<sup>34</sup>. Previous analysis on TCGA pan-cancer database involving 10,225 samples from 32 cancers also revealed that *TP53* mutations were associated with poorer prognosis in human cancer<sup>35</sup>. In this study, *TP53* mutations were associated with poorer prognosis of AFPGC patients using Kaplan–Meier analysis (Supplementary Fig. 11), but not an independent prognostic factor. CUB and sushi multiple domains (*CSMD3*) was shown to participate in dendrite development<sup>36</sup>. Previous studies have reported that *CSMD3* is involved in tumorigenesis and tumor proliferation in lung cancer<sup>37</sup> and liver cancer<sup>38</sup>. Moreover, *CSMD3* acts as an immune regulator and might be associated with immune response<sup>39,40</sup>. The formyl peptide receptor 1 (*FPR1*) is one of the members of the formyl peptide receptor family, which is involved in cell motility, angiogenesis, inflammation, and immune response<sup>41–43</sup>. *FPR1* showed association with epithelial mesenchymal transition, proliferation, and migration in different cancer types<sup>44,45</sup>. In this study, *FPR1* tended to show mutations more frequently in AFPGC than TCGA-GC or TCGA-CIN. Moreover, *FPR1*-mutated-AFPGC patients tended to have a higher distant metastatic rate when compared to *FPR1*-wild-type patients (80% vs. 38%,  $P = 0.149$ ). A recent study demonstrated that *FPR1* played a key role in chemotherapy-induced anticancer immune response<sup>40</sup>. In our analysis, we found that the frequency of *KMT2C* mutations in AFPGC was significantly higher than that in TCGA-CIN. Previous studies reported that Lysine (K)-specific methyltransferase 2C (*KMT2C*), encoding an H3K4 histone methyltransferase, was crucial for tumorigenesis and progression in various cancer<sup>46,47</sup>. Notably, a recent study demonstrated that *KMT2C* mutations promoted epithelial-to-mesenchymal transition in gastric adenocarcinoma<sup>48</sup>. Apart from significantly mutated genes, we also compared the frequently mutated genes (>10%) between AFPGC and TCGA-GC/TCGA-CIN (Supplementary Data 9 and 10). Understanding the role of key mutations in tumorigenesis,



tumor progression, and metastasis are important areas of AFPGC research in the future.

SCNA is defined as DNA segments (~1 kb or larger) that showed copy number differences among haplotypes<sup>49</sup>. SCNAs were responsible for evolution, genetic diversity, and genomic disorders including cancers<sup>50,51</sup>. So far, previous reports on SCNA analysis of AFPGC were limited. Wang et al.<sup>52</sup> have

reported that the frequent copy number gains at 20q11.21-13.12 in hepatoid adenocarcinoma of the stomach (a differentiated type of AFPGC) tended to be related to more adverse biobehaviour, including poorer differentiation, greater vascular and nerve invasion, and greater liver metastasis, but showed no statistically significant differences. In this study, we detected frequent copy number amplification alterations containing, e.g., *CCNE1*

**Fig. 5 Validation of selected genes as potential therapeutic targets in PDX models of AFPGC.** **a** The status of *AFP*, *ERBB2*, and *CCNE1* in AFPGC PDX models. The CNVs status of *ERBB2* and *CCNE1* are defined by the WES results of corresponding patients. **b–d** The representatives of AFPGC PDX models with different status of *ERBB2* and *CCNE1*. Scale bar represents 50  $\mu\text{m}$ . Data were obtained from three independent experiments. The right panel represents the CNVs status of *ERBB2* and *CCNE1* in the corresponding patients. **e–h** In vivo sensitivity of AFPGC with or without *CCNE1* amplification to *CCNE1* inhibition (AZD5438) ( $n = 5$  biologically independent samples). Tumor volumes and proportion of tumor growth inhibition were shown as means  $\pm$  SD. **i–k** In vivo sensitivity of *ERBB2*-positive AFPGC with or without *CCNE1* amplification to *ERBB2* (Trastuzumab) and *CCNE1* (AZD5438) dual inhibition ( $n = 5$  biologically independent samples). Tumor volumes and proportion of tumor growth inhibition were shown as means  $\pm$  SD. **l** The western blot analysis of critical molecules (*ERBB2*, ERK, p-ERK, AKT, p-AKT, CDK2, RB, and p-RB) for *ERBB2* and *CCNE1* co-positive PDX models (PDX8) were performed. The western blotting images are representative of three different independent experiments. Statistical significance was determined using two-sided Student's *t*-test (**e–g**) and one-way ANOVA (**i–k**). biw, twice a week; Chr, chromosome; CNVs, copy number variations; FISH, fluorescence in situ hybridization; HE, hematoxylin–eosin staining; IHC, immunohistochemistry; i.p., intraperitoneal injection; kD, kilodalton; p.o., oral gavage; qd, per day. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ; n.s., not significant. Source data are provided as a Source Data file.

(19q12), *ERBB2* (17q12), *CCND1* (11q13.3), *MUC4* (3q29), and *MCL1* (1q21.3).

In this study, we found that AFPGC presented with recurrent CNAs. Considering chromosomal instability (CIN) is a hallmark for TCGA-CIN<sup>8,53</sup>, we, therefore, compared the significant SCNA profiles between AFPGC and TCGA-CIN. Nearly 57% statistically significant amplification regions and 73% deletion regions were found in AFPGC specifically. These SCNA regions in AFPGC contain important cancer-related genes, such as *TRIM28* (19q13.43), which has been proved to accelerate cell proliferation and metastasis in a variety of human cancer<sup>54</sup>, and *FAT1* (1p36.32), which was observed in multiple human cancers to promote Wnt/ $\beta$ -catenin signaling and tumorigenesis<sup>55</sup>. CIN is a common feature of gastrointestinal adenocarcinomas (GIACs)<sup>8,53,56</sup> and nearly 70% of GIACs can be classified into CIN subtype<sup>57</sup>. Recent studies further classified CIN subtype of GC as well as other GIAC into subclasses with different genomic alterations and clinical significance<sup>57,58</sup>, suggesting that CIN subtype of GC may still be a heterogeneous group. A recent study also demonstrated that CIN is a driver of metastatic progression<sup>59</sup>. This might partially contribute to the aggressive phenotype of AFPGC. In addition, Bakhroum et al. observed that metastasis from CIN-high tumor cells tended to involved multiple organs<sup>59</sup>. Similarly, we also found that CIN subtype of GC displayed a multi-organ metastatic pattern by analyzing an MSK cohort reported by Janjigian et al.<sup>60</sup>. However, AFPGC showed a dominant tendency of liver metastasis (Supplementary Fig. 12), suggesting the different metastatic patterns between AFPGC and CIN subtype of GC. In all, the role of CIN in AFPGC is interesting and worth further exploration in larger cohorts by multi-omics in the future.

Combined with WES and clinical information, AFPGC patients with amplifications at 17q12 or 19q12 showed significantly worse survival than those without amplification. It is noteworthy that *ERBB2* and *CCNE1* have been proposed as two well-known oncogenes at 17q12 and 19q12, respectively<sup>61–63</sup>. *ERBB2* acts as a critical therapeutic target and targeting *ERBB2* in the treatment of *ERBB2*-positive metastatic GC has proven to be an effective therapeutic strategy<sup>64</sup>. *CCNE1* amplification is associated with poorer survival rate in patients with different types of cancers<sup>65,66</sup> and has been reported to associate with liver metastasis in *TP53*-mutated GCs<sup>67</sup>. The frequency of *ERBB2* and *CCNE1* amplification were 32.4% and 26.7% in our large cohort of AFPGC. Also, we demonstrated that both *ERBB2* and *CCNE1* amplification showed significant correlation with poor OS and several aggressive clinicopathological characteristics of AFPGC. This study reported that patients who harbored both *ERBB2* and *CCNE1* amplification demonstrated the worst prognosis in AFPGC. Furthermore, multivariate Cox regression analysis demonstrated that a positive status of *ERBB2* or *CCNE1* was an independent prognostic factor in AFPGC; however, this prognostic value was not observed in TCGA-CIN. ACRG-MSS/TP53<sup>−</sup> is another subtype of GC classification reported by Cristescu

et al.<sup>9</sup> and enriches *ERBB2* and *CCNE1* amplification. The prognostic value of *ERBB2* and *CCNE1* amplification in ACRG-MSS/TP53<sup>−</sup> was distinct from that in AFPGC (Supplementary Fig. 13). These findings implied the important role of *ERBB2* and *CCNE1* amplification in the development and progression of AFPGC.

To further explore the translational significance of the above findings, a series of AFPGC PDX models were established in our laboratory. Previous studies have shown that PDXs could well maintain the principal histological and genetical characteristics of their donor tumors and remain stable across the passages<sup>68,69</sup>. Our research group has previously applied GC PDX as a platform for evaluating drug sensitivity<sup>70,71</sup>. In this study, a large collection of AFPGC PDX models was reported. *ERBB2* and *CCNE1* as two frequently altered genes were taken as a translational example to explore the precise therapeutic target for this disease. Cyclin E1 (*CCNE1*), along with its catalytic subunit CDK2, plays a critical role in regulating the G1–S phase transition<sup>72</sup>. It has been reported that *CCNE1* amplification acts as a positive marker of CDK2 inhibitor sensitivity in breast as well as ovarian cancers<sup>18,73</sup>. AZD5438, a CDK1/2/9 inhibitor, exerts a significant antitumor effect in previous studies<sup>17,74</sup>. In our cohort, nearly 30% of AFPGCs harbored *CCNE1* amplification. Our *in vivo* study revealed that AZD5438 showed significant antitumor activity in *CCNE1*-amplified AFPGC PDX models. In our study, AFPGC patients involved a higher proportion of *ERBB2*-positive cases (32.4%) when compared with common GC (12–20%)<sup>75</sup>. Also, the *ERBB2*-positive GC PDX showed a good response to trastuzumab, which was consistent with the previous studies<sup>64,76</sup>. In our study, nearly 20% of *ERBB2*-positive AFPGC was accompanied by *CCNE1* amplification. Interestingly, our data showed that the co-positive of *ERBB2* and *CCNE1* were correlated with more aggressive behavior and poorer prognosis than single positive or dual negative. Previous studies have shown that the co-amplification of *CCNE1* and *ERBB2* is associated with trastuzumab resistance in breast cancer and GC<sup>64,76</sup>. Moreover, it has been reported that CDK2 inhibitor could augment the anti-proliferative impact of lapatinib (*ERBB2* inhibitor) in *CCNE1/ERBB2* co-amplified GC cells<sup>77</sup>. Therefore, we hypothesized that the dual-targeted therapy combination of *ERBB2* and *CCNE1* for AFPGC could be an effective therapeutic strategy to overcome or reverse the potential treatment resistance and have a synergistic antitumor effect. Encouragingly, AZD5438 was able to augment the antitumor effects of trastuzumab in AFPGC PDX model with co-amplification of *ERBB2/CCNE1*. Our results initially suggested that dual inhibition of *ERBB2* and *CCNE1* acts as a potential treatment strategy for AFPGCs with *ERBB2/CCNE1* co-amplification. Besides *CCNE1* and *ERBB2*, our study also provided a list of other potential therapeutic targets, deserving further exploration and verification. However, clinical trials investigating the above potential targets are unlikely to occur.

PDX model in our study plays an important platform for translational research and animal clinical trials for AFPGC.

There are some limitations that require acknowledgement in our study. First, in order to efficiently select AFPGCs from more than 5200 GC patients, we set relatively strict inclusion criteria, which may rise to selection bias to some extent. Although this study represented large genomic analyses of AFPGC samples, our analyses were still limited by sample size. That is, the results, such as significantly mutated genes, chromosomal alterations, and the correlation between genetic alterations and clinicopathological variables (e.g., histopathology), should still be fully validated in a larger multi-centered cohort. Because of the retrospective study design, the sample quality could not meet the criteria of multiple omics sequencing. Moreover, limited by the delayed approval of trastuzumab by the National Medical Products Administration of China and relatively low affordability of Chinese patients, the proportion of patients who received trastuzumab treatment in this study was relatively low, leading to insufficient cases to evaluate the efficacy of ERBB2-targeted therapy. PDX was an effective technology to explore therapeutic strategies. The effect of targeting therapeutic genomic alterations was evaluated in limited AFPGC PDXs due to scarcity of the disease and low success rate to generate and recover PDXs. More PDX samples can enhance the clinical implication of this study. Thus, a clinically reliable platform with more AFPGC PDX models should be developed and more clinical trials should be conducted in the future.

In summary, our study presented a large genomic landscape of AFPGC so far, which is a critical step in our efforts for understanding this disease. Besides, a series of PDX models were generated, which could be a platform to identify potential molecular targets and optimize therapeutic approaches for AFPGC.

## Methods

**Patients and samples.** From January 2011 to December 2018, there were 121 cases of primary gastric carcinoma with elevated serum AFP levels ( $\geq 20$  ng/ml) at diagnosis among 5261 GC cases at the First Affiliated Hospital, School of Medicine, Zhejiang University. One hundred and five patients with AFP-positive IHC staining were finally enrolled. For comparison, we randomly selected 1:3 cases (311 patients) with stage-matched primary GCs with normal serum AFP levels in our institution. We retrospectively collected fresh-frozen or FFPE tumor tissues and matched tumor adjacent normal tissues from 58 AFPGC patients for genomic characterization, after excluding 47 patients with prior chemotherapy (23 patients), insufficient tumor volume (12 patients), and low quality of DNA (12 patients). Staging was performed according to the American Joint Committee on Cancer TNM Staging Classification for Carcinoma of the Stomach (8th edition, 2018). Follow-up data were obtained by phone, letter, and the out-patient clinical database (last follow-up was September 2019) and follow-up information were available in 101 (101/105, 96.2%) patients. The OS time was calculated from the date of diagnosis to the last day of follow-up or the date of death. Patient-derived paraffin-embedded tissue samples were used in accordance with ethical guidelines in the First Affiliated Hospital, School of Medicine, Zhejiang University (No. 2018-309). Patient-written consents were obtained from study participants, informing the use for genomic sequencing, drug test, and publication.

**DNA extraction and DNA quantification and qualification.** For fresh-frozen samples, genomic DNA from tumor tissues and matched normal tissues were isolated using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. Further, for FFPE tumor tissues, DNA was extracted using QIAamp DNA FFPE Tissue Kit (Qiagen). Then we combined the following two methods to verify the quality of isolated genomic DNA. First, DNA degradation and contamination were monitored on 1% agarose gels. Second, Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Invitrogen, USA) was used to quantify DNA concentration.

**WES and data analysis.** Whole-exome library construction was generated using the Agilent SureSelect Human All Exon V6 Kit (Agilent Technologies, Santa Clara, CA, USA). According to the manufacturer's instructions, the index-coded samples were clustered on a cBot Cluster Generation System using HiSeq PE Cluster Kit (Illumina). After cluster generation, the DNA libraries were sequenced on Illumina HiSeq platform (Illumina, San Diego, California, USA) and 150 bp paired-end reads were generated.

Data quality control was first performed and all downstream bioinformatics analyses were based on high-quality clean data, in which reads containing an adapter, reads containing poly-N, and low-quality reads were removed. The paired-end clean reads were aligned to the Human Genome Reference Consortium build 37 (GRCh37) using BWA v.0.7.8<sup>78</sup>.

### Somatic mutation detection and significantly mutated genes identification.

Identification of somatic SNVs was conducted by muTect<sup>79</sup> and the somatic InDels were detected by Strelka<sup>80</sup>. Paired adjacent normal tissue DNA was obtained as a control for all tumor samples. To reduce false-positive calls in FFPE specimens, only mutation site with a minimum of three variant reads and a variant allele frequency  $> 0.08$  was used for further analysis. ANNOVAR (ANNOVAR\_2015Mar22)<sup>81</sup> was used to annotate variant call format files. The MuSiC algorithm<sup>13</sup> was used to identify significantly mutated genes from the profiles of somatic SNVs and InDels in AFPGC (false discovery rate (FDR)  $< 0.25$ ).

**Mutational signatures analysis.** There are six variant types of single base substitution as follows: C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G, and T>G/A>C. According to the number of point mutations of different types, we conducted a cluster analysis to observe the similarity and difference within tumor samples. Considering the type of base at 1 bp in the upstream and downstream of the point mutations, the point mutation was divided into 96 types. According to the frequency of 96 mutation types, the point mutation types were decomposed into several different mutation characteristics by non-negative matrix factorization method<sup>82</sup>. Then, extracted mutational signatures were compared to the pan-cancer catalog of 30 known signatures referenced in the Catalogue of Somatic Mutations in Cancer (COSMIC) database using SomaticSignatures packages<sup>83</sup>. The similarity of mutation signatures was evaluated with cosine similarity  $> 0.85$ , which suggests common signatures. Signatures 1, 17, and 29 were identified in our samples. There were two signatures common to signature 1 and we kept the higher value of cosine similarity signature as signature 1.

**Copy number analysis.** We identified SCNAs using CNVkit<sup>84</sup> to analyze the copy number state of each tumor. Then we used GISTIC 2.0<sup>85</sup> to evaluate the genome regions with significant amplification or deletion in the samples and screen out the regions with high frequencies, namely recurrent CNA regions. The GISTIC score reflects the frequency of CNA in this segment. Red and blue indicate an increase and a decrease in the number of copies, respectively. Then we downloaded TCGA-GC copy number data from TCGA and identified significant SCNAs using GISTIC, to compare AFPGC SCNA profiles with TCGA-GC and subtypes (TCGA-CIN, TCGA-EBV, TCGA-GS, and TCGA-MSI).

**Integrated pathway analysis.** We downloaded the geneset of cancer-driving genes from Cancer Gene Census in the COSMIC, then investigated the genetic alteration rate of the cancer-driving genes in AFPGC. Subsequently, we screened out the frequently altered genes and performed KEGG pathway enrichment using DAVID bioinformatics Resource 6.8. We found several key pathways significantly enriched in AFPGC (FDR  $< 0.05$ ). Besides, we downloaded TCGA-GC data from TCGA and compared genetic alterations (somatic mutations and SCNAs) of the key pathways in AFPGC and TCGA-GC.

**Analysis of potential targetable genes.** We screened out a list of genes altered at least in four cases in our AFPGC cohort that are potentially targetable (based on COSMIC, OncoKB databases<sup>86</sup>, MSK-IMPACT<sup>87</sup>, and comprehensive literature review). These genetic alterations of targetable genes were compared between AFPGC and TCGA-GC, then visualized by ComplexHeatmap package<sup>88</sup>.

**H&E and IHC staining.** Tumor tissues were fixed in 10% neutral buffered formalin followed by gradient dehydration, wax immersion, embedding, and sectioning, then evaluated by hematoxylin and eosin (H&E) and IHC staining. For H&E staining, sections were processed by dewaxing, hydration, dyeing, dehydration, and sealing. For IHC staining, sections were dewaxed, rehydrated, and subjected to antigen retrieval. Then we incubated the sections with primary antibodies against ERBB2 (1:200, Cell Signaling Technology, #2165) and AFP (1:100, ProteinTech, #14550-1-AP) at 4 °C for 12 h after quenching endogenous peroxidase activity and blocking nonspecific binding sites. This was followed by a 30 min incubation with secondary antibody [SP Rabbit & Mouse HRP Kit (DAB) (1:400, CoWin Biosciences, #CW2069)]. We performed IHC using the streptavidin-biotin peroxidase complex method (Lab Vision, Fremont, CA, USA), then observed and photographed the sections using an optical microscope (Nikon, Tokyo Japan). IHC results were evaluated according to previously published methods<sup>64,89</sup> and details were shown in Supplementary Methods. ERBB2-positive status included ERBB2 overexpression (score 3+) by IHC or ERBB2 amplification by FISH as described in previous studies<sup>64,90</sup>.

**Fluorescence in situ hybridization.** FISH analysis was performed for ERBB2 and CCNE1 gene assessment on FFPE tumor tissues using ERBB2/CEN17 dual color Probe (ZytoVision, #Z-2020-20) and CCNE1/Con19 FISH Probe (Empire

Genomics, #CCNE1-CHR19-20-ORGR) following the manufacturers' recommendations. Probes were co-denatured for 10 min at 75 °C (ZytoVision)/3 min at 83 °C (Empire Genomics) on the slide and incubated overnight at 37 °C. Three post-hybridization washes were performed in 1 × Wash Buffer A at 37 °C for 5 min each. Finally, the slides were air dried and counterstained with 4',6-diamidino-2-phenylindole/antifade solution. Signals for each locus-specific FISH probe were evaluated under an Olympus BX51 microscope (Olympus Corporation, Tokyo, Japan). CCNE1 gene amplification results were based on at least 50 evaluable tumor nuclei. CCNE1 amplification was defined as the presence of either loose or tight CCNE1 cluster, or the ratio of CCNE1/Con19  $\geq 2.0$ <sup>91</sup>. ERBB2 gene status was classified as non-amplified (ERBB2/CEN17 ratio < 2.0) or amplified (ERBB2/CEN17 ratio  $\geq 2.0$ )<sup>64,92</sup>. Once an adequate target area had been identified, the scores for ERBB2 and CEN17 copy numbers present in 20 representative nuclei were scored. If the resulting ratio of ERBB2/CEN17 fell within 1.8–2.2, an additional 20 nuclei were scored and the resulting ratio was calculated from a total of 40 nuclei.

**Western blotting.** Lysates of frozen tumor tissue were extracted by lysis buffer containing protease and phosphatase inhibitors (Sigma-Aldrich, St. Louis, MO, USA), then fractionated by SDS-polyacrylamide gel electrophoresis and electro-transferred to polyvinylidene difluoride membranes (Millipore, Billerica, MA, USA). After blocking in 5% skim milk in triethanolamine buffered saline-tween for 1 h at room temperature, the membranes were incubated with primary antibodies overnight at 4 °C. The membranes were washed and then incubated with peroxidase-conjugated secondary antibodies [HRP Conjugated AffiniPure Goat Anti-rabbit IgG (H + L) (1 : 5000, BOSTER, #BA1054) and HRP Conjugated AffiniPure Goat Anti-mouse IgG (H + L) (1 : 5000, BOSTER, #BA1051)] for 1 h at room temperature. After washing three times with triethanolamine buffered saline, the blot was soaked in ECL™ chemiluminescent detection reagents (Millipore, Billerica, MA, USA) for 1 min. The membranes were then exposed to film (Kodak Rochester, NY, USA) for 30 s in a darkroom. The antibodies against the following proteins were used: ERBB2 (1 : 1000, Cell Signal, #2165), AKT (1 : 1000, Cell Signal, #4691), p-AKT (1 : 1000, Cell Signal, #13038), ERK (1 : 1000, Cell Signal, #4695), p-ERK (1 : 2000, Cell Signal, #4370), CDK2 (1 : 3000, ProteinTech, 10122-1-AP), Rb (1 : 1000, ProteinTech, 17218-1-AP), p-Rb (1 : 1000, Cell Signal, #8516), and  $\beta$ -Actin (1 : 500, BOSTER, #BM0627).

**Establishing xenografts and treatment protocol.** Four- to six-week-old female BALB/c nude mice, purchased from Shanghai Slac Laboratory Animal Corporation (Shanghai, China), were housed with regular 12 h light/12 h dark cycles for at least 3 days before use. Ambient temperature was 20–22 °C, kept at constant humidity of 40–60%. PDX models were established as we previously reported<sup>71,93</sup>. We monitored xenograft growth at least twice weekly by a vernier caliper measuring the length (*L*) and width (*W*) of the tumor, and then removed them for serial transplantation after the volume reached about 1500 mm<sup>3</sup>. The tumor volume (*V*) was calculated according to the following formula:  $V = L \times W^2/2$ .

We used xenografts from the third generation (the second mouse-to-mouse passage) for the experiments once the tumor volume reached about 150–200 mm<sup>3</sup>. Mice were randomly assigned to different groups (five to seven mice/group) as follows: (i) vehicle; (ii) Trastuzumab (SelleckChem) 10 mg/kg twice weekly of intraperitoneal injection; (iii) AZD5438 (SelleckChem) 20 mg/kg daily by oral gavage; (iv) AZD5438 + Trastuzumab, for 3 weeks. Experiments were ended once the tumor volume surpassed 1500 mm<sup>3</sup> or mouse weight loss reached 20%. The percentage of tumor growth inhibition (TGI) was calculated according to the following formula:  $TGI = (1 - T/C) \times 100\%$ , where *T/C* represents the relative tumor volume of treatment group and control group. After the mice had been killed, we conducted immunoblot to assess the expression of various markers. Animal care and experiments were performed under the approval and supervision of the Animal Experimental Ethical Inspection of the First Affiliated Hospital, College of Medicine, Zhejiang University (number 2018-378).

**Statistical analysis.** Two or multiple continuous variables were compared using Student's *t*-test or one-way analysis of variance, respectively. The  $\chi^2$  and Fisher's exact tests were used to analyze the significance of categorical data. Survival rates were calculated by the Kaplan–Meier method and compared using the log-rank test. Multivariate Cox regression analysis was performed to identify the independent prognostic factors. Non-parametric variables were compared using the Kruskal–Wallis test and the Mann–Whitney test. Statistical analysis was performed using SPSS 21.0 software. *P* < 0.05 was considered statistically significant.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The raw sequence data reported in this study have been deposited in the Genome Sequence Archive<sup>94</sup> in National Genomics Data Center<sup>95</sup>, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences [<http://bigd.big.ac.cn/gsa-human/>] (accession number: HRA000429). The whole-exome somatic variants are also publicly available from the European Nucleotide Archive [<https://www.ebi.ac.uk/ena/browser/home>] (study ID: PRJEB44858 (ERP128950)).

The clinical data are provided in Supplementary Data 1. A complete list of TCGA cohort can be found in Supplementary Data 11. The MSKCC data are available in the cBioPortal for Cancer Genomics database [[https://www.cbioportal.org/study/summary?id=egc\\_msk\\_2017](https://www.cbioportal.org/study/summary?id=egc_msk_2017)]. The TCGA data are available in the cBioPortal for Cancer Genomics database [[https://www.cbioportal.org/study/summary?id=stad\\_tcga](https://www.cbioportal.org/study/summary?id=stad_tcga)]. TCGA data analyzed for this manuscript were released on 28 January 2016. All other data are available within the Article and Supplementary Information. Source data are provided with this paper.

Received: 21 August 2020; Accepted: 1 June 2021;

Published online: 24 June 2021

## References

- Hirasaki, S. et al. Seronegative alpha-fetoprotein-producing early gastric cancer treated with endoscopic mucosal resection and additional surgery. *Intern. Med.* **43**, 926–930 (2004).
- Gong, W. et al. Clinical characteristics and treatments of patients with alpha-fetoprotein producing gastric carcinoma. *Neoplasma* **65**, 326–330 (2018).
- Nagtegaal, I. D. et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 182–188 (2020).
- Hirajima, S. et al. Liver metastasis is the only independent prognostic factor in AFP-producing gastric cancer. *World J. Gastroenterol.* **19**, 6055–6061 (2013).
- Liu, X. et al. Clinicopathologic features and prognostic factors in alpha-fetoprotein-producing gastric cancers: analysis of 104 cases. *J. Surg. Oncol.* **102**, 249–255 (2010).
- Kono, K. et al. Clinicopathologic features of gastric cancers producing alpha-fetoprotein. *Dig. Surg.* **19**, 359–365 (2002). *Discussion* 365.
- Chang, Y. C. et al. Comparison between the clinicopathologic features of AFP-positive and AFP-negative gastric cancers. *Am. J. Gastroenterol.* **87**, 321–325 (1992).
- Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
- Cristescu, R. et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **21**, 449–456 (2015).
- Ge, S. et al. A proteomic landscape of diffuse-type gastric cancer. *Nat. Commun.* **9**, 1012 (2018).
- Mun, D. G. et al. Proteogenomic characterization of human early-onset gastric cancer. *Cancer Cell* **35**, 111–124 e110 (2019).
- Gullo, I. et al. The transcriptomic landscape of gastric cancer: insights into Epstein-Barr virus infected and microsatellite unstable tumors. *Int. J. Mol. Sci.* **19**, <https://doi.org/10.3390/ijms19072079> (2018).
- Dees, N. D. et al. MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
- Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Christensen, S. et al. 5-Fluorouracil treatment induces characteristic T>G mutations in human cancer. *Nat. Commun.* **10**, 4571 (2019).
- Au-Yeung, G. et al. Selective targeting of cyclin E1-amplified high-grade serous ovarian cancer by cyclin-dependent kinase 2 and AKT inhibition. *Clin. Cancer Res.* **23**, 1862–1874 (2017).
- Chen, Z. et al. Characterization and validation of potential therapeutic targets based on the molecular signature of patient-derived xenografts in gastric cancer. *J. Hematol. Oncol.* **11**, 20 (2018).
- Scaltriti, M. et al. Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients. *Proc. Natl Acad. Sci. USA* **108**, 3761–3766 (2011).
- Alpert, E., Pinn, V. W. & Isselbacher, K. J. Alpha-fetoprotein in a patient with gastric carcinoma metastatic to the liver. *N. Engl. J. Med.* **285**, 1058–1059 (1971).
- Ishikura, H. et al. Hepatoid adenocarcinomas of the stomach. An analysis of seven cases. *Cancer* **58**, 119–126 (1986).
- Kinjo, T. et al. Histologic and immunohistochemical analyses of alpha-fetoprotein-producing cancer of the stomach. *Am. J. Surg. Pathol.* **36**, 56–65 (2012).
- Kodama, T. et al. Production of alpha-fetoprotein, normal serum proteins, and human chorionic gonadotropin in stomach cancer: histologic and immunohistochemical analyses of 35 cases. *Cancer* **48**, 1647–1655 (1981).
- Matsunou, H., Konishi, F., Jalal, R. E., Yamamichi, N. & Mukawa, A. Alpha-fetoprotein-producing gastric carcinoma with enteroblastic differentiation. *Cancer* **73**, 534–540 (1994).
- Motoyama, T., Aizawa, K., Watanabe, H., Fukase, M. & Saito, K. alpha-Fetoprotein producing gastric carcinomas: a comparative study of three different subtypes. *Acta Pathol. Jpn* **43**, 654–661 (1993).

25. Uefuji, K., Ichikura, T. & Tamakuma, S. Roles of histological findings and serum AFP levels in the prognosis of AFP-producing gastric cancers. *Jpn J. Clin. Oncol.* **24**, 135–140 (1994).
26. Liu, X., Sheng, W. & Wang, Y. An analysis of clinicopathological features and prognosis by comparing hepatoid adenocarcinoma of the stomach with AFP-producing gastric cancer. *J. Surg. Oncol.* **106**, 299–303 (2012).
27. Japanese Gastric Cancer, A. Japanese gastric cancer treatment guidelines 2018 (5th edition). *Gastric Cancer* **24**, 1–21 (2021).
28. Lee, J. Y. et al. The impact of concomitant genomic alterations on treatment outcome for trastuzumab therapy in HER2-positive gastric cancer. *Sci. Rep.* **5**, 9289 (2015).
29. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
30. Arora, K. et al. Fetal-type gastrointestinal adenocarcinoma: a morphologically distinct entity with unfavourable prognosis. *J. Clin. Pathol.* **71**, 221–227 (2018).
31. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **174**, 1034–1035 (2018).
32. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
33. Sabapathy, K. & Lane, D. P. Therapeutic targeting of p53: all mutants are equal, but some mutants are more equal than others. *Nat. Rev. Clin. Oncol.* **15**, 13–30 (2018).
34. Bullock, A. N., Henckel, J. & Fersht, A. R. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene* **19**, 1245–1256 (2000).
35. Donehower, L. A. et al. Integrated analysis of TP53 gene and pathway alterations in The Cancer Genome Atlas. *Cell Rep.* **28**, 1370–1384 e1375 (2019).
36. Shimizu, A. et al. A novel giant gene CSMD3 encoding a protein with CUB and sushi multiple domains: a candidate gene for benign adult familial myoclonic epilepsy on human chromosome 8q23.3–q24.1. *Biochem. Biophys. Res. Commun.* **309**, 143–154 (2003).
37. Liu, P. et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* **33**, 1270–1276 (2012).
38. Lai, M. W. et al. Hepatocarcinogenesis in transgenic mice carrying hepatitis B virus pre-S/S gene with the sW172\* mutation. *Oncogenesis* **5**, e273 (2016).
39. Kentsis, A. et al. Urine proteomics for discovery of improved diagnostic markers of Kawasaki disease. *EMBO Mol. Med.* **5**, 210–220 (2013).
40. Ahn, J. W. et al. Identification of somatic mutations in EGFR/KRAS/ALK-negative lung adenocarcinoma in never-smokers. *Genome Med.* **6**, 18 (2014).
41. Chen, K. et al. Formylpeptide receptor-2 contributes to colonic epithelial homeostasis, inflammation, and tumorigenesis. *J. Clin. Invest.* **123**, 1694–1704 (2013).
42. McCoy, R. et al. N-formylpeptide and complement C5a receptors are expressed in liver cells and mediate hepatic acute phase gene regulation. *J. Exp. Med.* **182**, 207–217 (1995).
43. Vacchelli, E. et al. Chemotherapy-induced antitumor immunity requires formyl peptide receptor 1. *Science* **350**, 972–978 (2015).
44. Prevede, N. et al. The formyl peptide receptor 1 exerts a tumor suppressor function in human gastric cancer by inhibiting angiogenesis. *Oncogene* **34**, 3826–3838 (2015).
45. Khau, T. et al. Annexin-1 signals mitogen-stimulated breast tumor cell proliferation by activation of the formyl peptide receptors (FPRs) 1 and 2. *FASEB J.* **25**, 483–496 (2011).
46. Gala, K. et al. KMT2C mediates the estrogen dependence of breast cancer through regulation of ERalpha enhancer function. *Oncogene* **37**, 4692–4710 (2018).
47. Shah, A. et al. Acquired cystic kidney disease-associated renal cell carcinoma (ACKD-RCC) harbor recurrent mutations in KMT2C and TSC2 genes. *Am. J. Surg. Pathol.* **44**, 1479–1486 (2020).
48. Cho, S. J. et al. KMT2C mutations in diffuse-type gastric adenocarcinoma promote epithelial-to-mesenchymal transition. *Clin. Cancer Res.* **24**, 6556–6569 (2018).
49. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
50. Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
51. Leary, R. J. et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl Acad. Sci. USA* **105**, 16224–16229 (2008).
52. Wang, Y. et al. Hepatoid adenocarcinoma of the stomach: a unique subgroup with distinct clinicopathological and molecular features. *Gastric Cancer* **22**, 1183–1192 (2019).
53. Cancer Genome Atlas Research, N. et al. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169–175 (2017).
54. Czerwinska, P., Mazurek, S. & Wiznerowicz, M. The complexity of TRIM28 contribution to cancer. *J. Biomed. Sci.* **24**, 63 (2017).
55. Morris, L. G. et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nat. Genet.* **45**, 253–261 (2013).
56. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
57. Liu, Y. et al. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **33**, 721–735 e728 (2018).
58. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* **20**, 404–416 (2019).
59. Bakhrouf, S. F. et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**, 467–472 (2018).
60. Janjigian, Y. Y. et al. Genetic predictors of response to systemic therapy in esophagogastric cancer. *Cancer Discov.* **8**, 49–58 (2018).
61. Bieche, I. et al. Two distinct amplified regions at 17q11–q21 involved in human primary breast cancer. *Cancer Res.* **56**, 3886–3890 (1996).
62. Lin, L. et al. Identification and characterization of a 19q12 amplicon in esophageal adenocarcinomas reveals cyclin E as the best candidate gene for this amplicon. *Cancer Res.* **60**, 7021–7027 (2000).
63. Natrajan, R. et al. Functional characterization of the 19q12 amplicon in grade III breast cancers. *Breast Cancer Res.* **14**, R53 (2012).
64. Bang, Y. J. et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* **376**, 687–697 (2010).
65. Zhao, H. et al. Prognostic values of CCNE1 amplification and overexpression in cancer patients: a systematic review and meta-analysis. *J. Cancer* **9**, 2397–2407 (2018).
66. Pils, D. et al. Cyclin E1 (CCNE1) as independent positive prognostic factor in advanced stage serous ovarian cancer patients - a study of the OVCAD consortium. *Eur. J. Cancer* **50**, 99–110 (2014).
67. Kim, B. et al. CCNE1 amplification is associated with liver metastasis in gastric carcinoma. *Pathol. Res. Pract.* **215**, 152434 (2019).
68. Tentler, J. J. et al. Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* **9**, 338–350 (2012).
69. Hidalgo, M. et al. Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014).
70. Jin, K. et al. Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review. *Clin. Transl. Oncol.* **12**, 473–480 (2010).
71. Wang, H. et al. Establishment of patient-derived gastric cancer xenografts: a useful tool for preclinical evaluation of targeted therapies involving alterations in HER-2, MET and FGFR2 signaling pathways. *BMC Cancer* **17**, 191 (2017).
72. Hwang, H. C. & Clurman, B. E. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24**, 2776–2786 (2005).
73. Kanska, J., Zakhour, M., Taylor-Harding, B., Karlan, B. Y. & Wiedemeyer, W. R. Cyclin E as a potential therapeutic target in high grade serous ovarian cancer. *Gynecol. Oncol.* **143**, 152–158 (2016).
74. Yang, L. et al. Cyclin-dependent kinase 2 is an ideal target for ovary tumors with elevated cyclin E1 expression. *Oncotarget* **6**, 20801–20812 (2015).
75. Van Cutsem, E. et al. HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer. *Gastric Cancer* **18**, 476–484 (2015).
76. Tanner, M. et al. Amplification of HER-2 in gastric carcinoma: association with Topoisomerase IIalpha gene amplification, intestinal type, poor prognosis and sensitivity to trastuzumab. *Ann. Oncol.* **16**, 273–278 (2005).
77. Kim, J. et al. Preexisting oncogenic events impact trastuzumab sensitivity in ERBB2-amplified gastroesophageal adenocarcinoma. *J. Clin. Invest.* **124**, 5145–5158 (2014).
78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
79. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
80. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
81. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
82. Tan, V. Y. & Fevotte, C. Automatic relevance determination in nonnegative matrix factorization with the beta-divergence. In *IEEE Trans. on Pattern Analysis and Machine Intelligence* **35**, 1592–1605 (IEEE, 2013).
83. Gehring, J. S., Fischer, B., Lawrence, M. & Huber, W. SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**, 3673–3675 (2015).

84. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
85. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
86. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precision Oncol.* **2017**, <https://doi.org/10.1200/PO.17.00011> (2017).
87. Cheng, D. T. et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagnostics* **17**, 251–264 (2015).
88. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
89. Murakami, T. et al. Clinicopathologic and immunohistochemical characteristics of gastric adenocarcinoma with enteroblastic differentiation: a study of 29 cases. *Gastric Cancer* **19**, 498–507 (2016).
90. Bartley, A. N. et al. HER2 Ttesting and clinical decision making in gastroesophageal adenocarcinoma: guideline from the College of American Pathologists, American Society for Clinical Pathology, and the American Society of Clinical Oncology. *J. Clin. Oncol.* **35**, 446–464 (2017).
91. Kuhn, E. et al. CCNE1 amplification and centrosome number abnormality in serous tubal intraepithelial carcinoma: further evidence supporting its role as a precursor of ovarian high-grade serous carcinoma. *Mod. Pathol.* **29**, 1254–1261 (2016).
92. Das, K. et al. Mutually exclusive FGFR2, HER2, and KRAS gene amplifications in gastric cancer revealed by multicolour FISH. *Cancer Lett.* **353**, 167–175 (2014).
93. Jin, K. et al. Establishment of a PDTT xenograft model of gastric carcinoma and its application in personalized therapeutic regimen selection. *Hepatogastroenterology* **58**, 1814–1822 (2011).
94. Wang, Y. et al. GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics* **15**, 14–18 (2017).
95. National Genomics Data Center, M. & Partners Database resources of the National Genomics Data Center in 2020. *Nucleic Acids Res.* **48**, D24–D33 (2020).

## Acknowledgements

We thank Zheng Ju and Hong Su (from Novogene Bioinformatics Technology Co. Ltd, Beijing, China) for assistance with bioinformatics analyses. We greatly thank all the staff and assistants in the Department of Surgical Oncology, the First Affiliated Hospital, Zhejiang University School of Medicine, for their support in collecting samples. This project was supported by National Natural Science Foundation of China (Grant number 81803107), Zhejiang Provincial Natural Science Foundation of China (LQ18H160002 and LQ20H160043), the Key Research and Development Program of Science and Technology Department of Zhejiang Province (2018C03022), Zhejiang Provincial Department of Science and Technology (number 2020C03112), and Project of the

Regional Diagnosis and Treatment Centre of the Health Planning Committee (number JBZX-201903).

## Author contributions

L.T., Haiyong Wang, Z.Z., J.L., Y.D., and Yiran Chen are responsible for the study concept and design. J.L., Y.D., J.J., and Yiran Chen are responsible for the development of methodology. Yanyan Chen, J.J., Yiran Chen, Y.H., M.W., C.L., and M.K. are responsible for the acquisition of data. J.L., Y.D., Yanyan Chen, W.Z., and Z.Z. are responsible for the analysis and interpretation of data. J.L., Y.D., Yanyan Chen, J.J., and Yiran Chen are responsible for the writing of the manuscript. H.Z., F.T., Haiyong Wang, Z.Z., and L.T. are responsible for the review and/or revision of the manuscript. Haohao Wang, J.Z., Z.L., Y.L., X.Y., and K.J. are responsible for the administrative, technical, or material support. D.Z., T.Z., Z.Z., Haiyong Wang, and L.T. are responsible for study supervision. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24170-0>.

**Correspondence** and requests for materials should be addressed to Z.Z., H.W. or L.T.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021