Supplementary Information

Circulating tumor DNA analsyes predict progressive disease and indicate trastuzumab-resistant mechanism in advanced gastric cancer

Authors: Yan Wang,[1†] Chuanhua Zhao, [1†] Lianpeng Chang,[2†] Ru Jia,[1] Rongrui Liu,[1] Yun Zhang,[1] Xuan Gao,[2] Jin Li,[2] Rongrong Chen,[2] Xuefeng Xia,[2] Ajaz Bulbul, [3] Hatim Husain, [3] Yanfang Guan,[2] Xin Yi,[2*] Jianming Xu[1*]

Supplementary Materials and Methods

Study design

The objectives of this prospective study were to develop a strategy for accurate measurement of tumor load with targeted capture sequencing of serial ctDNA and explain mechanisms of drug resistance in AGC. We validated ctDNA clonal mutations derived from tumor tissue and calculated the mTBI with the mean clonal mutation fraction in cfDNA. We applied mTBI to assess 21 patients with metastatic AGC receiving trastuzumab. The analysis provided coverage of 100% of the patients and a mean lead time in detection of progressive disease of 18 weeks before imaging result detection. We identified expanding mutation clones during treatment and evaluated their relationship with resistance to trastuzumab. A validation patient group revealed the feasibility of using mTBI in patients receiving chemotherapy. We also assessed whether mTBI could serve as a predictive marker of treatment outcome.

Patients and samples

Patients with AGC were drawn from the Fifth Medical Center, General Hospital of PLA, Beijing, China. Serial peripheral blood (10 mL) was sampled from each patient at baseline and every 2–4 treatment cycles. Fourteen matched formalin-fixed, paraffin-embedded (FFPE) tissues were obtained whenever sufficient slides were available. All tumor samples were obtained from primary biopsies and had ≥60% tumor cell content as determined by histopathology assessment.

Peripheral blood was collected in EDTA Vacutainer tubes (BD Diagnostics, Franklin Lakes, NJ, USA) and processed within 2 h to separate plasma and buffy coat (as a source samples of germline DNA). Plasma was separated by centrifugation at 1,600 g for 10 min, transferred to microcentrifuge tubes, and centrifuged at 16,000 g for 10 min to remove remaining cell debris. Buffy coat from the first centrifugation were used for the extraction of germline genomic DNA. Both plasma and buffy coat were stored at −80°C until the time of DNA extraction.

Circulating DNA was isolated from plasma using a QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany). Buffy coat DNA and tumor tissue DNA were extracted using the DNeasy Blood & Tissue Kit (Qiagen). DNA concentration was measured using a Qubit fluorometer and the Qubit dsDNA HS (High

Sensitivity) Assay Kit (Invitrogen, Carlsbad, CA, USA). The size distribution of the cfDNA was assessed using an Agilent 2100 BioAnalyzer and a DNA HS kit (Agilent Technologies, Santa Clara, CA, USA).

Pan-cancer panel sequencing

Before library construction, 1 μg each of tissue and buffy coat DNA was sheared to 300-bp fragments with a Covaris S2 ultrasonicator. Indexed Illumina NGS libraries were prepared from tissue, and germ line and circulating DNA libraries were prepared using the KAPA Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA). Target enrichment was performed with a custom SeqCap EZ Library (Roche NimbleGen, Madison, WI, USA). Capture hybridization was carried out according to the manufacturer's protocol. Two versions of capture probe were used to enrich libraries. Version 1.0 (~2.3 M) applied in samples from patients received chemotherapy plus trastuzumab, was designed to cover coding sequence of 386 frequently mutated gene in solid tumors. Version 2.0 (~1.1 M) applied in samples from patients receiving chemotherapy only, were designed to cover hot exons and hot regions (1021 genes) frequently mutated in solid tumors. Genes and coordinates of selected regions of each version are provided in Table S1. Following hybrid selection, the captured DNA fragments were amplified and then pooled to generate several multiplex libraries.

Sequencing was carried out using Illumina 2×100 bp paired-end reads on an Illumina HiSeq 3000 instrument according to the manufacturer's recommendations using a TruSeq PE Cluster Generation Kit v3 and a TruSeq SBS Kit v3 (Illumina, San Diego, CA, USA).

Identification of somatic mutation in tumor tissue and ctDNA

Terminal adaptor sequences and low-quality reads were removed from raw data of paired samples. Burrows-Wheeler Aligner (BWA; version 0.7.12-r1039) was employed to align the clean reads to the reference human genome (hg19). Picard (version 1.98) was used to mark polymerase chain reaction (PCR) duplicates. Realignment and recalibration was performed using GATK (version 3.4-46-gbc02625). Single nucleotide variants (SNV) were called using MuTect (version 1.1.4) and NChot [1], a software developed in-house to review hotspot variants. Small insertions and deletions (InDels) were determined by GATK. Somatic copy number alterations were identified with CONTRA (v2.0.8). Significant copy number

variation was expressed as the ratio of adjusted depth between circulating DNA and germline DNA. The candidate variants were all manually verified in the Integrative Genomics Viewer (IGV).

Mutations were considered as a candidate somatic mutation only when (i) the mutation had at least five high-quality reads (Phred score ≥30, mapping quality ≥30, and without paired-end reads bias) containing the particular base; (ii) the mutation was not presented in >1% of population in the 1,000 Genomes Project or dbSNP databases (The Single Nucleotide Polymorphism Database); and (iii) the mutation was not present in a local database of normal samples. For tumor somatic mutations, a mutant allele must be present in ≥2% of reads. Non-synonymous mutations annotated by ANNOVAR were used in clonal structure reconstruction.

Clone structure and mTBI analysis

PyClone was employed to analyze the clonal structure, based on a Bayesian clustering method. Depth of normal and variant allele was generated using Sequence Alignment/Map tools (SAMtools), including reads with Phred score ≥30 and mapping quality ≥30. Candidate mutations in different ctDNA of the same patient were retrieved whenever variant alleles read ≥3. Copy number was generated from Contra. The independent input was used to analyze clonal structure in ctDNA and tissue for ctDNA at baseline and matched tissue samples, respectively. For serial ctDNA, multi-input of each sample was used to analyze serial clonal population. For each clustering process, a PyClone algorithm was run for 20,000 iterations with a burn-in of 2,000 iterations using a beta binomial model with the "total_copy_number" option. The maximum VAFs of somatic mutations in each sample were used as "Tumor content." Other parameters were default [2]. Cancer cell fraction was calculated with the mean of predicted cellular frequencies. The cluster with highest mean VAF was identified to be clonal cluster, and mutations in this cluster were clonal mutations. Meanwhile, other clusters and mutations were consider to be subclonal. In each ctDNA sample, mTBI was analyzed with the mean VAF of clonal mutations. The ΔmTBI was calculated based on mTBI of the first ctDNA sample.

Pathway analysis of expanding clones

Mutations detected at PD ctDNA with absolute increased VAF ≥2 % or with the increased CCF≥2-fold compared with that of BL ctDNA, were considered to be high-confidence expanding clones. SIFT and

PolyPhen2 were used for predicting a functional impact of an amino acid substitution caused by mutations. WebGestalt carried out pathway enrichment analysis to investigate the distribution of genes affected by somatic mutations and CNVs within the KEGG database [3].

Statistical analysis

Pearson correlation analysis was used to test the linear association between CNV and ΔmTBI. Multivariate Cox proportional hazards analysis (enter method) was performed considering clinical characteristics and mTBI at baseline. Kaplan-Meier survival plots were generated for mTBI at baseline using log-rank tests. All statistical analyses were performed with SPSS (v.21.0; STATA, College Station, TX, USA) or GraphPad Prism (v. 6.0; GraphPad Software, La Jolla, CA, USA) software. Statistical significance was defined as a two-sided *P*-value of <0.05.

Supplementary references

1. Yang X, Chu Y, Zhang R, Han Y, Zhang L, Fu Y, et al. Technical Validation of a Next-Generation Sequencing Assay for Detecting Clinically Relevant Levels of Breast Cancer–Related Single-Nucleotide Variants and Copy Number Variants Using Simulated Cell-Free DNA. The Journal of Molecular Diagnostics 2017.

2. Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun 2015;6:8760.

3. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res 2017.

Supplementary Figures

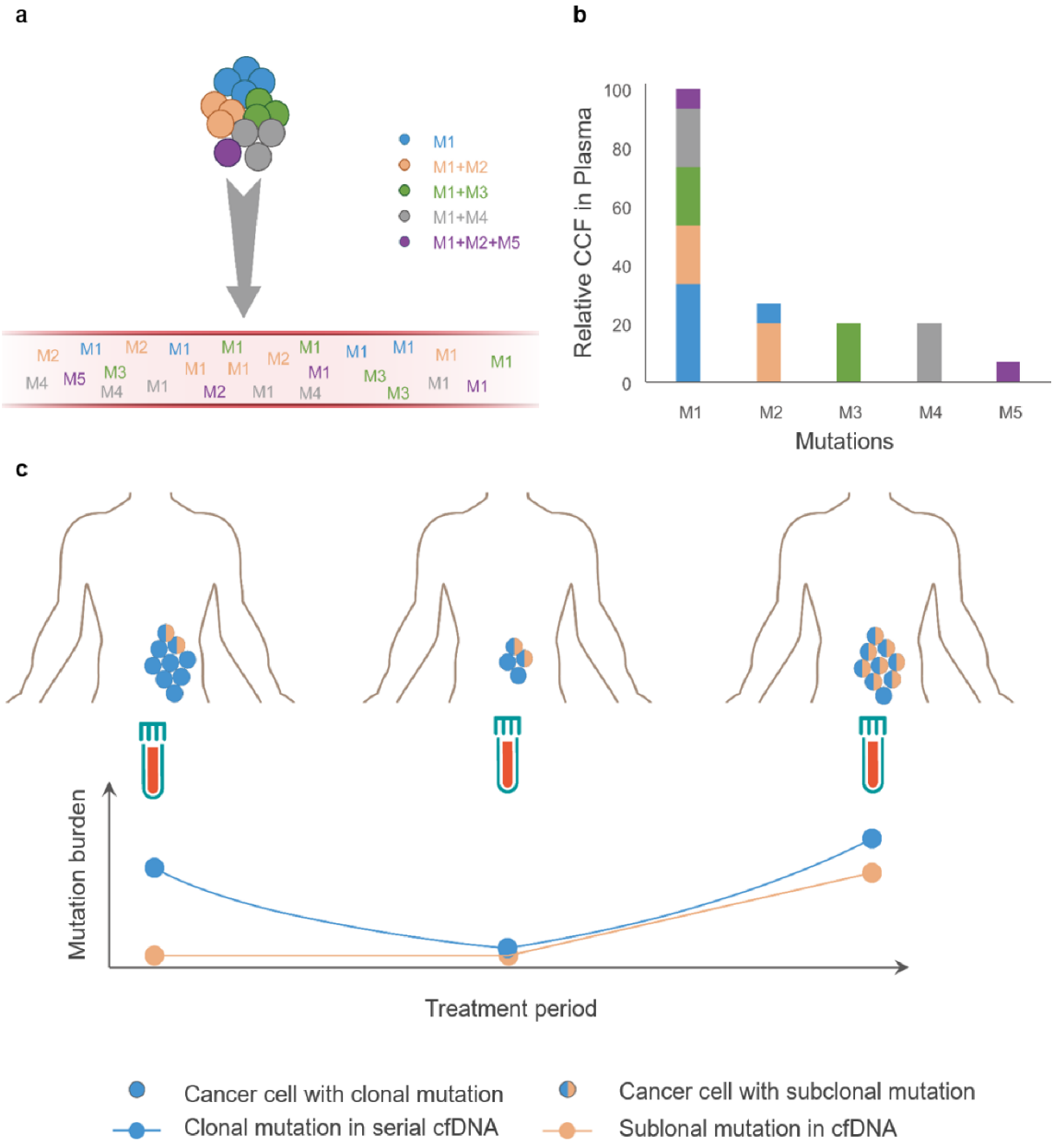

Fig. S1. Schematic diagram of clone analysis and dynamic change in ctDNA. (a) Clonal mutations with the higher cancer cell fraction (CCF) in tissue present higher abundance of ctDNA. (b) Targeted capture sequencing analysis of the relative CCF of each mutation from tissue cells and identified the clonal mutation. (c) During treatment, clonal mutations in ctDNA changed with tumor burden, and expanding subclonal mutations were potentially related to the drug resistance mechanism.

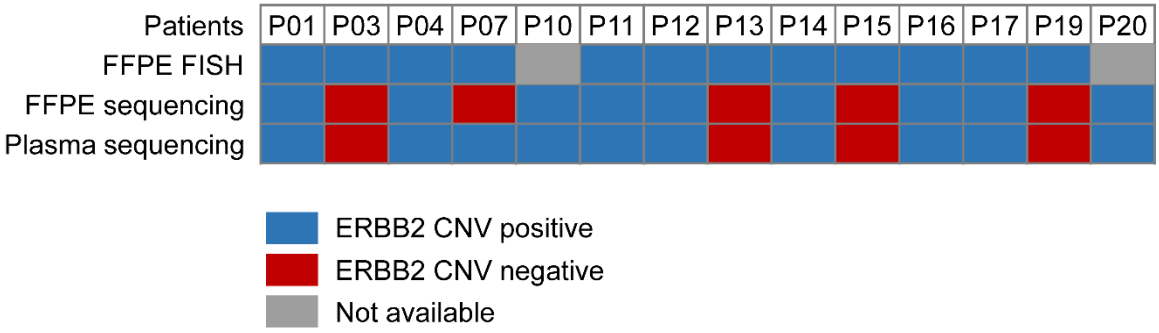

Fig. S2. Comparison of CNV in ERBB2 gene using FISH and pan-cancer panel sequencing in paired samples. FFPE samples of P10 and P20 were both IHC 3+ and not detected using FISH.

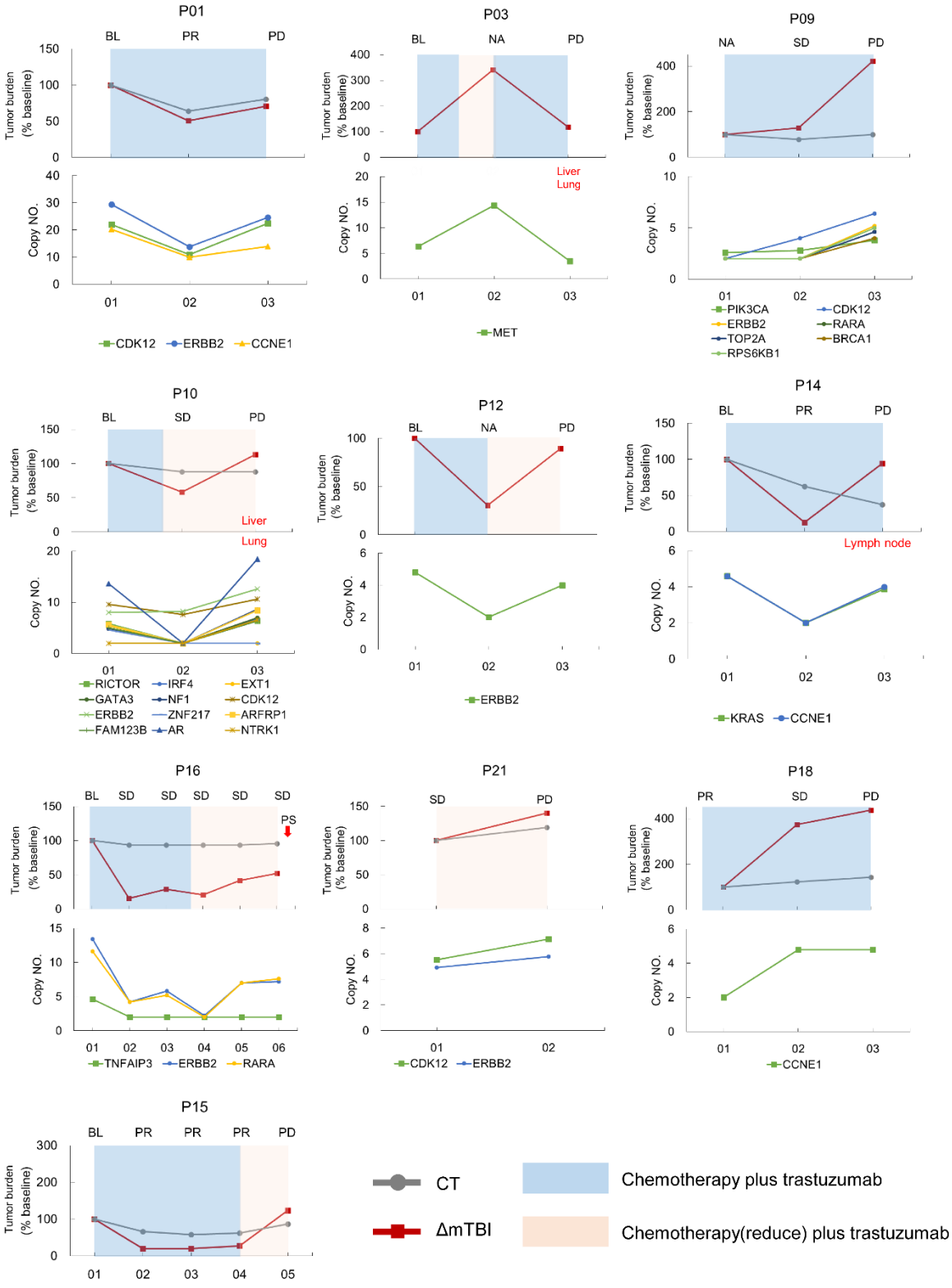

Fig. S3. Serial changes of mTBI in patients observed progressive disease during anti-Her2 treatment. Top: Tumor burden changed based on the first observed result. ΔmTBI (red line) was calculated based on the

first sample. The gray line indicated CT imaging results. Progressive metastasis were marked as red words at PD. Bottom: serial changes of CNV detected in ctDNA. BL, baseline; PR, partial response; PD, progressive disease; NA, no available CT result; PS, palliative surgery; mTBI, molecular tumor burden index; CT, computed tomography.

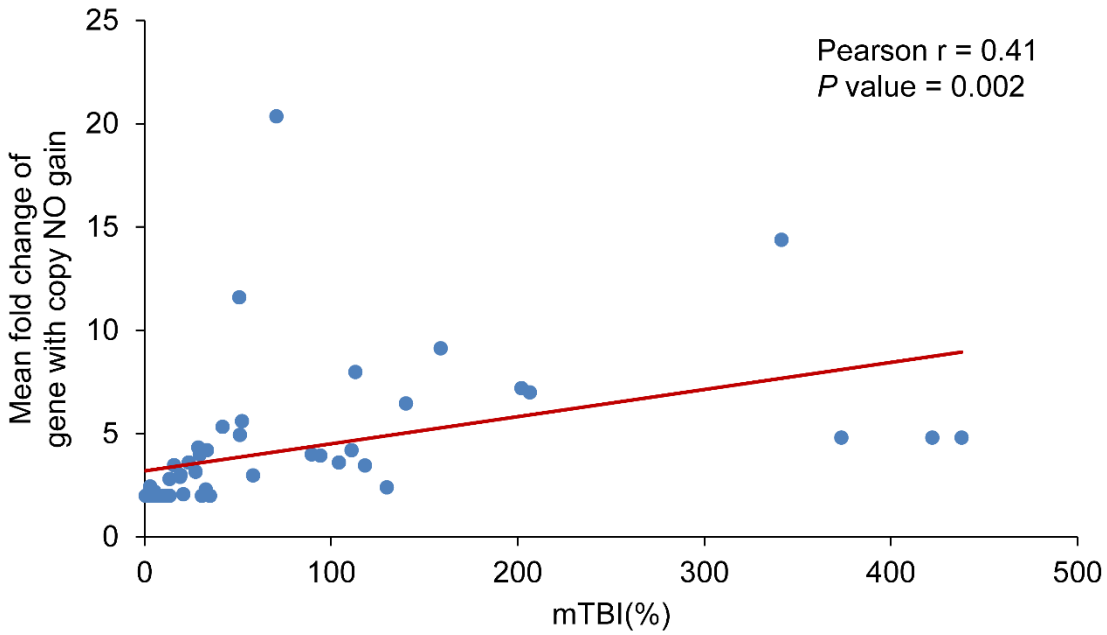

Fig. S4. Relation between mTBI and average copy NO. of CNV. The status of mTBI and CNV were statistically significant (Pearson r=0.41, *P*=0.002), which enhanced the reliability of mTBI for serial monitoring.

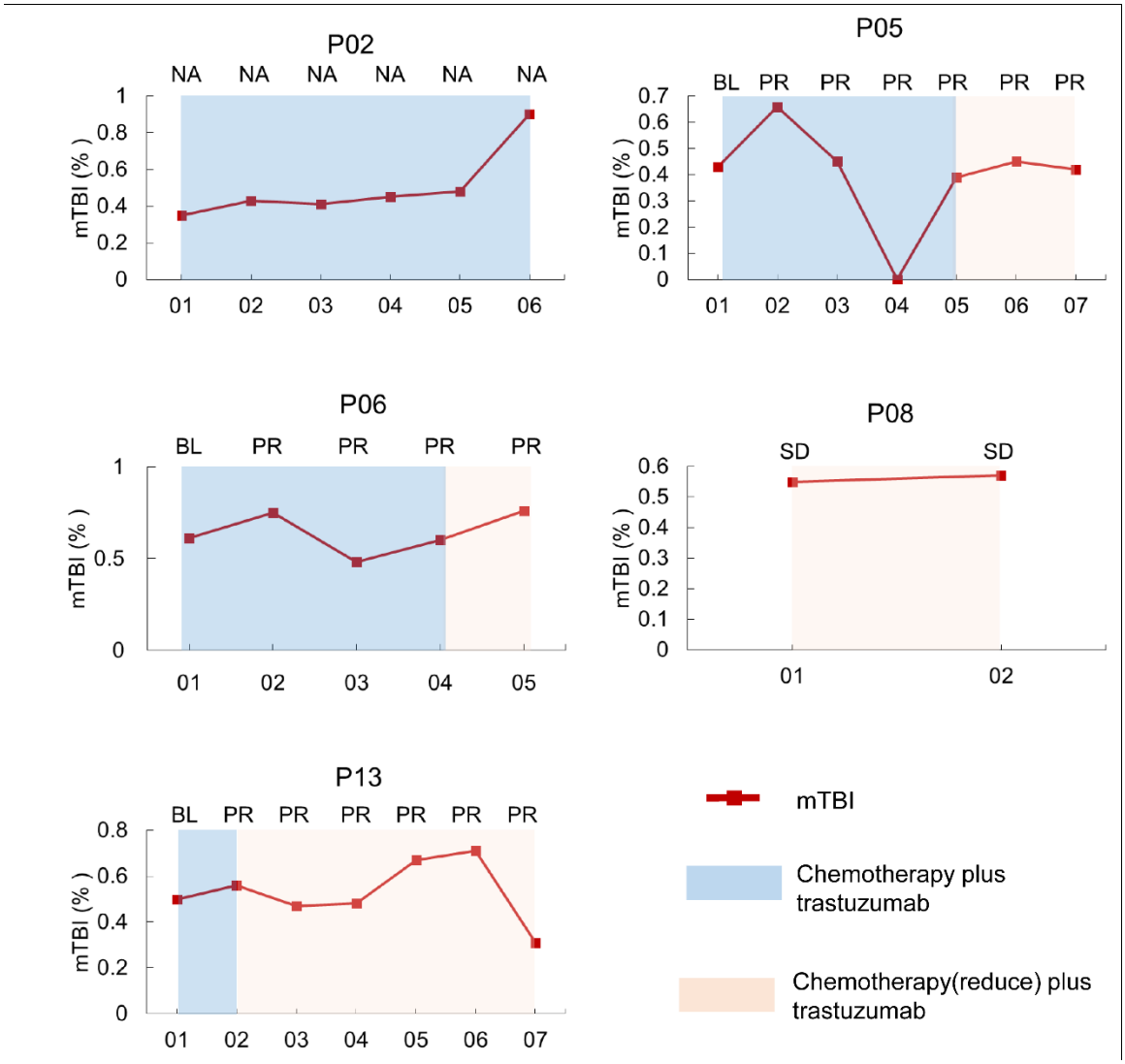

Fig. S5. Serial changes of mTBI in patients without progressive disease during anti-Her2 treatment. NA, no available CT result; BL, baseline; PR, partial response; PD, progressive disease; mTBI, molecular tumor burden index; CT, computed tomography.

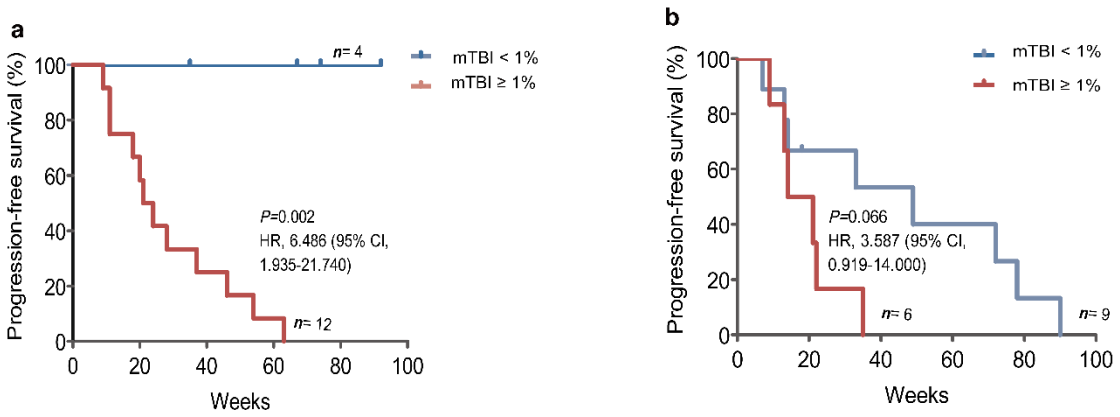

Fig. S6. Kaplan-Meier analysis of progression-free survival in patients treated with chemotherapy plus trastuzumab and with chemotherapy alone. Kaplan-Meier analysis showed patients with high (≥1%) baseline mTBI presented shorter PFS than patients with lower (<1%) mTBI in both chemotherapy plus trastuzumab treated AGC (a) and chemotherapy treated AGC (b) (chemotherapy plus trastuzumab subgroup, $P = 0.002$; chemotherapy subgroup, $P = 0.066$).

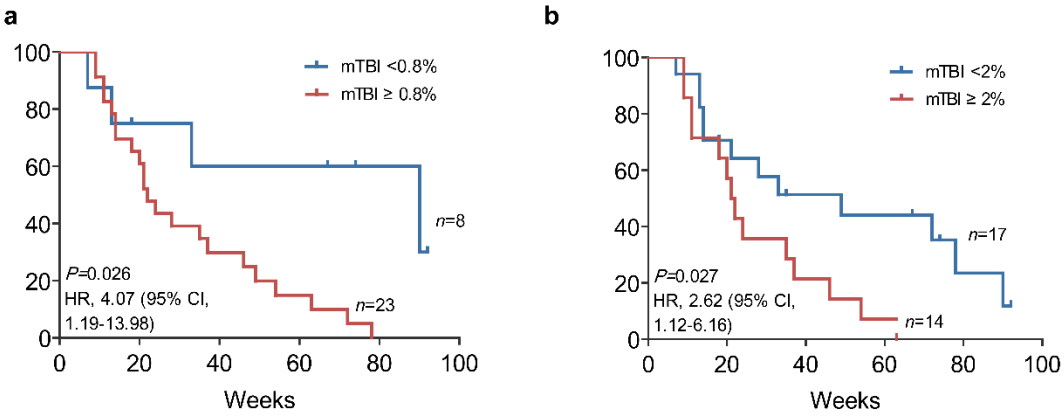

Fig. S7. Kaplan-Meier analysis of progression-free survival in patients with high or low mTBI. Different cutoff values were used in this analysis: 0.8 % (a) and 2 % (b). Both results indicated high mTBI at baseline was associated with worse outcome (pretreatment mTBI of 0.8%, *P*=0.026; pretreatment mTBI of 2%, *P*=0.027).