



The Clinical and Molecular Characterization of Gastric Cancer Patients in Qinghai-Tibetan Plateau

Guanghong Rong^{1†}, Yongxia Zhang^{2†}, Yingcai Ma¹, Shilong Chen^{3,4} and Yongcui Wang^{4,5*}

¹ Department of Gastroenterology, Qinghai Provincial People's Hospital, Xining, China, ² Department of Gynecology, Qinghai Provincial People's Hospital, Xining, China, ³ Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China, ⁴ Institute of Sanjiangyuan National Park, Chinese Academy of Sciences, Xining, China, ⁵ Qinghai Provincial Key Laboratory of Crop Molecular Breeding, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China

OPEN ACCESS

Edited by:

Masanori Terashima,
Shizuoka Cancer Center, Japan

Reviewed by:

Lei Zhang,
Sun Yat-sen University, China
Xinliang Lu,
Zhejiang University, China

*Correspondence:

Yongcui Wang
ycwang@nwipb.cas.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Gastrointestinal Cancers,
a section of the journal
Frontiers in Oncology

Received: 08 April 2020

Accepted: 26 May 2020

Published: 30 June 2020

Citation:

Rong G, Zhang Y, Ma Y, Chen S and
Wang Y (2020) The Clinical and
Molecular Characterization of Gastric
Cancer Patients in Qinghai-Tibetan
Plateau. *Front. Oncol.* 10:1033.
doi: 10.3389/fonc.2020.01033

Gastric cancer was the fifth most common malignancy and the third deadliest cancer (738,000 deaths in 2018) in the world. The analysis of its molecular characteristics has been complicated by histological and intratumor heterogeneity. Furthermore, the previous studies indicate that the incidence of gastric cancer shows wide geographical variation. As the largest and highest region in China, Qinghai-Tibetan Plateau (QTP) is one of the important global biodiversity hotspots. Here, to better understand the mechanism of gastric cancer and offer the targeted therapeutic strategies specially designed for patients in QTP, we collect tumor and blood samples from 30 primary gastric adenocarcinoma cancer patients at Qinghai Provincial People's Hospital. We discuss the clinical and molecular characteristics for these patients that have been ascribed to the unique features in this place, including high altitude (the average height above sea level is around 4,000 m), multi-ethnic groups, and the specific ways of life or habits (such as eating too much beef and mutton, have alcohol and cigarette problem, et al.). By comparing with the western gastric cancer patients collected from TCGA data portal, some unique characteristics for patients in QTP are suggested. They include high incidence in younger people, most of tumor are located in body, most of SNP are detected in chromosome 7, and the very different molecular atlas in minor ethnic groups and Han Chinese. These characteristics will provide the unprecedented opportunity to increase the efficacy for diagnosis and prognosis of gastric cancer in QTP. Furthermore, to suggest the targeted therapeutics specially designed for these 30 patients, an adapted kernel-based learning model and a compilation of pharmacogenomics data of 462 patient-derived tumor cells (PDCs) that illustrate the diverse genetic and molecular backgrounds of cancer patients, were introduced. In conclusion, our study offers a big opportunity to better understand the mechanism of gastric cancer in QTP and guide the optimal patient-tailored therapy for them.

Keywords: gastric cancer, qinghai-tibetan plateau, clinical characteristics, molecular characteristics, the mechanism of cancer, cancer targeted therapeutics

INTRODUCTION

Although it is steadily declining in incidence, cancer of the stomach (also known as gastric cancer) remains one of the most common and deadly malignancies in the world (1, 2). According to GLOBOCAN 2018 data, gastric cancer is the fifth most common malignancy and the third deadliest cancer (738,000 deaths annually), following only lung and colorectal cancer in overall mortality. Based on the reports from cancer.net in 2020, the 5-years survival rate is around 69%, if the cancer is diagnosed and treated before it has spread outside the stomach, will drop to 31%, if the cancer has spread to surrounding tissues or organs and/or the regional lymph nodes, and is 5%, if the cancer has spread to a distant part of the body.

Gastric cancer incidence and mortality were used to be related with *Helicobacter pylori* (*H. pylori*) infection. Recently, strides in preventing and treating *H. pylori* infection have already made the incidence declining. However, gastric cancer remains the fifth highest incidence among cancers, with 5.7% of all new cases attributable to the disease (3). A better understanding of the mechanism and risk factors of the disease can enable earlier diagnosis and thus lead to greater survival. Dietary modification, smoking cessation, and exercise hold promise in preventing gastric cancer. Furthermore, the previous studies suggested that the incidence of gastric cancer shows wide geographical variation (3, 4). More than 50% of the new cases occur in developing countries. East Asia (China and Japan), Eastern Europe, Central and South America are area with high occurrence, and Southern Asia, North and East Africa, North America, Australia, and New Zealand are area with low occurrence (4).

As the world's largest developing country with divergent geographical variation, China is characterized by large disparities in health care. Qinghai-Tibetan Plateau (QTP), the largest and highest region in China, is one of the important global biodiversity hotspots. QTP has a complex geological history, and it is a common understanding that the central plateau uplifted first and formed the 'proto-QTP' as early as 40 Mya, followed by outward extensions in early Miocene (5–7). The late Miocene, 8–10 Mya, all the mountains surrounding QTP had reached steady elevation (8–11). Qinghai province is in northeast QTP. The average 4,000-m height above the sea level lead to the unique climate features there, including average temperature of 21 and –14 degrees centigrade in summer and winter, respectively, low humidity et al. People living there have their special dietary habit (such as eating too much beef and mutton), and always have alcohol and cigarette problem. All these features would result in unique characteristics of gastric cancer there, and should follow up the specific treatment strategy. In addition, according to the reports from Qinghai Provincial People's Hospital, the incidence of gastric cancer was the highest (640 new cases in 2019), followed by liver (427 new cases in 2019), lung (303 new cases in 2019), and colon cancer (237 new cases in 2019). Thus, it is urgent to analysis of its molecular and clinical characteristics for better understanding the mechanism of disease and guiding the optimal patient-tailored therapy.

Here, for better understanding the mechanism of gastric cancer and offer the targeted therapeutic strategies specially

designed for patients in QTP, we collect tumor and blood sample from 30 primary gastric adenocarcinoma cancer patients at Qinghai Provincial People's Hospital, and discuss their unique clinical and molecular characteristics that have been ascribed to the unique features in this area. By comparing with the western gastric cancer patients collected from TCGA data portal (12), some unique characteristics for patients in QTP are suggested. They include high incidence in younger people, most of tumor are located in body, most of SNP are in chromosome 7, and the very different driver genes from that in western patients. The further analysis on driver genes indicates the different molecular atlas in minor ethnic groups and Han Chinese. Using an adapted kernel-based learning model and a compilation of pharmacogenomics data of 462 patient-derived tumor cells (PDCs) that illustrate the diverse genetic and molecular backgrounds of cancer patients (13), the specific targeted therapeutics for these 30 patients are suggested.

MATERIALS AND METHODS

Tumor Specimens and Blood Sample

The tumor specimens and matched germline DNA samples (blood samples) with corresponding clinical records were obtained from patients undergoing surgery or gastroscopy with polypectomy at Qinghai Provincial People's Hospital. This work was performed in compliance with all relevant ethical regulations for research using human specimens. The fresh frozen tissue and blood samples were delivered to sequencing companies (ANOROAD GENOME and Sangon Biotech, **Table S1**) to capture the exonic DNA fragments and perform the whole exome sequencing. All patients here were acknowledged and signed the agreement with the hospital.

The Pharmacogenomics Data that Used for Training Drug Response Learning Mode

The data used to build the drug response learning model came from the drug response data on the patient-derived cancer cells, which illustrate the diverse genetic and molecular backgrounds of cancer patients. It includes pharmacological landscapes of 462 patient-derived cancer cells across 60 small molecules, together with genomic and transcriptomic profiling in 385 of these tumors (13). The somatic mutations for 91 genes were used as the cancer genomics and the compound chemical structure was used to represent drug.

TCGA Gastric Adenocarcinoma Cancer Patients and TCGA Drug Response Data

The clinical information for gastric adenocarcinoma cancer patients in TCGA data portal was collected (12). These information was used to compare with our plateau patients. The molecular properties for cancer patients, including somatic mutation and mRNA expression were collected from TCGA data portal to evaluate the prediction model in revealing the effective drugs for cancer patients. The clinical drug responses to TCGA cancer patients came from Ding's work (14). As did in (14), we removed possible combination therapy and chemotherapy earlier

than surgery, and kept 2,182 patient-drug pairs (including 1,029 pan-cancer patients and 130 clinical drug) for evaluation.

Cancer Cell Line Pharmacogenomics

To illustrate the advantage of using drug sensitivities in patient-derived cancer cells in prediction of the effective drugs for cancer patients, the well-established cancer genomic resources CTRP (15) was introduced. CTRP links genetic, lineage, and other cellular features of cancer cell lines to small-molecule sensitivity, and extended 24 compounds in Cancer Cell Line Encyclopedia (CCLE) (16) to 'Informer Set' compounds (including 481 small molecule probes and drugs). Specifically, the mutation status for 25 oncogenes across 486 cancer cells, and mRNA expressions for 54,675 mRNAs across 127 cancer cells were obtained from CCLE. The dose-response curve (AUCDR) for 545 drugs across 887 cancer cell lines were extracted from CTRP.

Whole Exome Sequencing and SNP Calling

The Illumina HiSeq 2000 instrument was used for sequencing and generated 2×150 base paired-end reads. FASTQ files were aligned to the human genome assembly (hg38) via Burrows-Wheeler Aligner (BWA) (17). Before further analysis, the initially aligned BAM files were subjected to preprocessing that sorted, removed duplicated reads, locally realigned reads around potential small indels, and recalibrated base quality scores using SAMtools (18), and Picard (<https://broadinstitute.github.io/picard/>). The single Nucleotide Polymorphism, SNP was then detected through the Genome Analysis ToolKit (GATK) (19). After removing the duplicated and low quality SNPs, the remaining SNPs were annotated via ANNOVAR (20).

Genome Coverage, Somatic Mutation, and Gene Fusion Analysis

The `genomecov` function under `bedtools` package (21) was used to calculate the coverage per-base based on preprocessed BAM files. Using the human genome (hg38) annotation file, the coverage at gene level was obtained. The somatic mutation was obtained by using germline SNPs from blood samples as a reference. The FusionMap (22) was introduced to perform the gene fusion analysis based on FASTQ files.

Identification of Cancer Driver Gene

The MaxMIF (23), which outperformed the existing state-of-the-art methods (including MUFFINN (24), MuttSig2 (25), MutSigCV (26), et al.) on TCGA pan-cancer datasets, was introduced to distinguish the cancer driver genes from the passenger genes. MaxMIF integrated the somatic mutation data and molecular interaction data by a maximal mutational impact function. The protein-protein interaction (PPI) network deposited in HumanNet (27) was introduced to represent the molecular interactions.

An Adapted Kernel-Based Learning Model for Prediction of the Effective Clinical Drugs

We applied our previously adapted kernel-based learning model to predict the effective clinical drugs for cancer patients (28).

To simultaneously consider all available drugs across all cancer cell lines, we introduced the bipartite graph framework under the assumption that drugs with similar chemical properties should have similar treatment outcomes (28). To overcome the difficulty of heterogeneous data integration, we adapted a kernel method to construct similarity matrix based on different types of feature. Specifically, the cancer genomic data (such as mutation, expression, et al.) and chemical properties were applied to construct kernel-based similarity matrices between cancer cell lines and anticancer drugs. Based on experimentally tested drug sensitivities, a bipartite graph was constructed to represent the relationships between cancer cells and anticancer drugs, where the edges were labeled by the digitalized drug sensitivities. The three classification models, random forest (RF), support vector machine (SVM), and deep learning network (DN) were then utilized on these kernel-based similarity matrices to predict the effective clinical drugs for cancer patients. The RF, SVM and DN model were implemented via "randomForest" R package with default parameters, LibSVM (29) in "e1071" R package with RBF kernel function, and the "h2o" R package with default parameters, respectively. The penalty parameter and the RBF kernel parameter were optimized by the grid search approach with 5-fold cross-validation. The area under the ROC curve (AUC) (30) was used as the evaluation criteria to assess the performance of classification model.

RESULTS

Clinical Characteristics

We obtained tumor tissue (fresh frozen) from 30 primary gastric adenocarcinoma cancer patients not treated with prior chemotherapy or radiotherapy (Table S1). We collected some clinical information for these cancer patients (Figure 1A), including the initial diagnosis age, the gender, the location of primary tumor tissue, and the ethnic group information, and compared them with western gastric adenocarcinoma patients which collected from TCGA (Figure 1B). Both western and plateau patients have higher incidence in male patients. The most tumors of plateau patients are located in body, while the antrum tumors are the most for western patients. Comparing with the western patients in terms of initial diagnosis age, plateau patients are younger. They are around fifty, and western patients are around sixty. The differences in clinical properties indicate the unique molecular characteristics for these plateau patients that differ with western patients, and should follow by the specific treatment strategy.

Molecular Characteristics

We first check the sequencing depth at gene level via counting the genome coverage along the gene loci. Most of genes have depth around a hundred (Figure 2A) and gene coverage shows the different distribution for different ethnic group of patients (Figure 2B). Then we discuss the SNP distribution along the chromosome and compare them with western patients. The SNP distribution along the chromosome is more stable in western patients than that in plateau patients (Figures 3A,C), and the compared with other chromosome, more SNPs were detected in

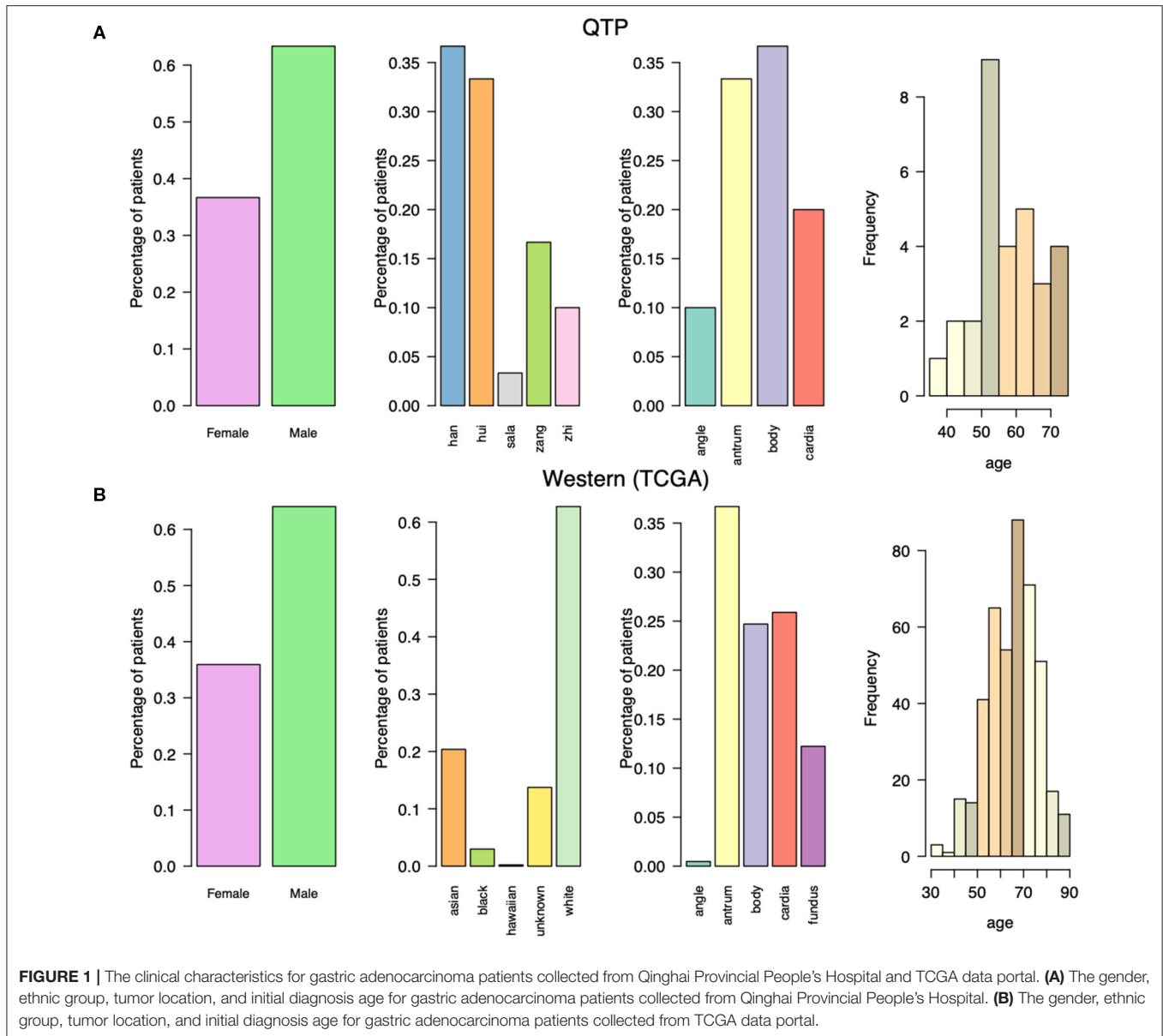
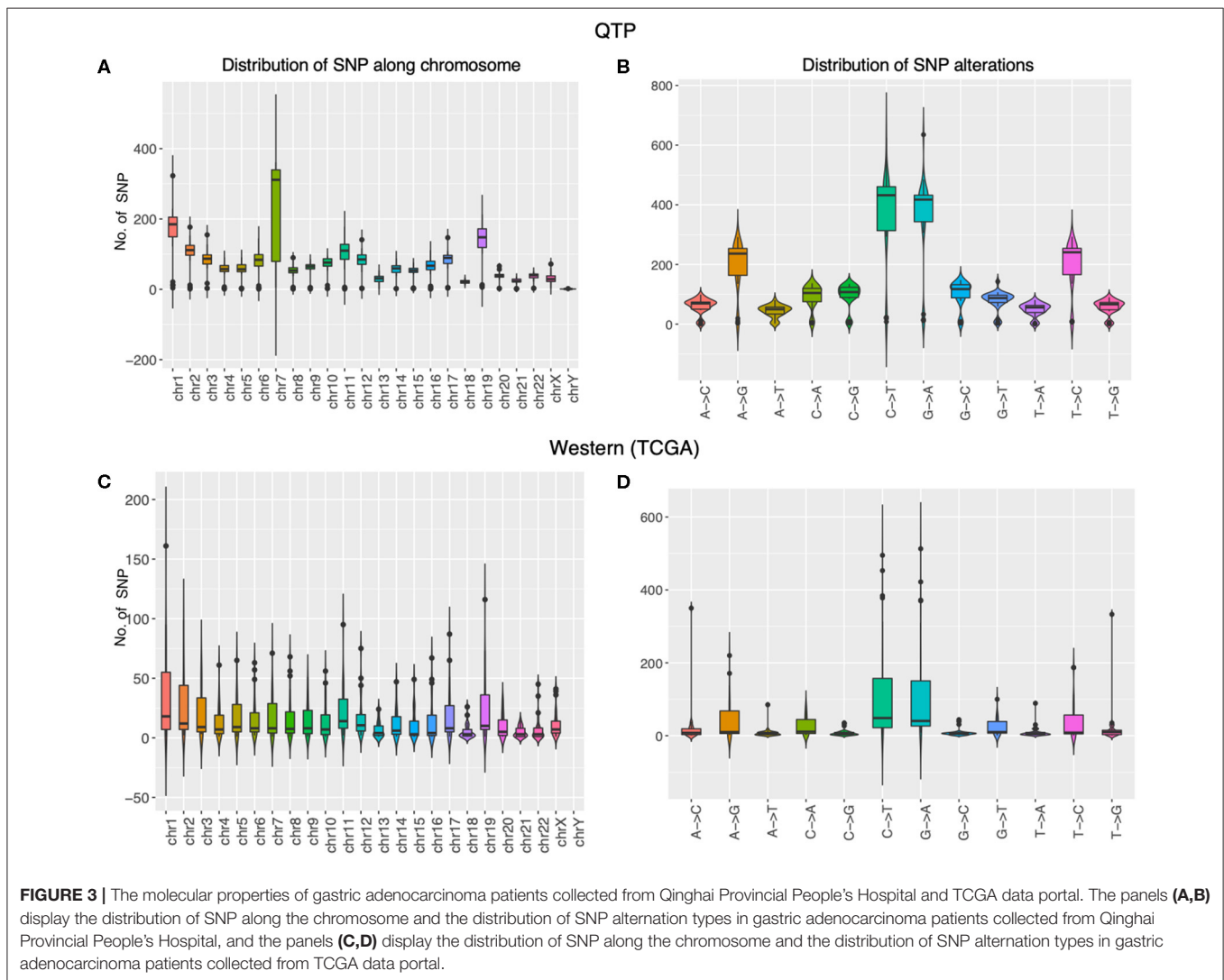
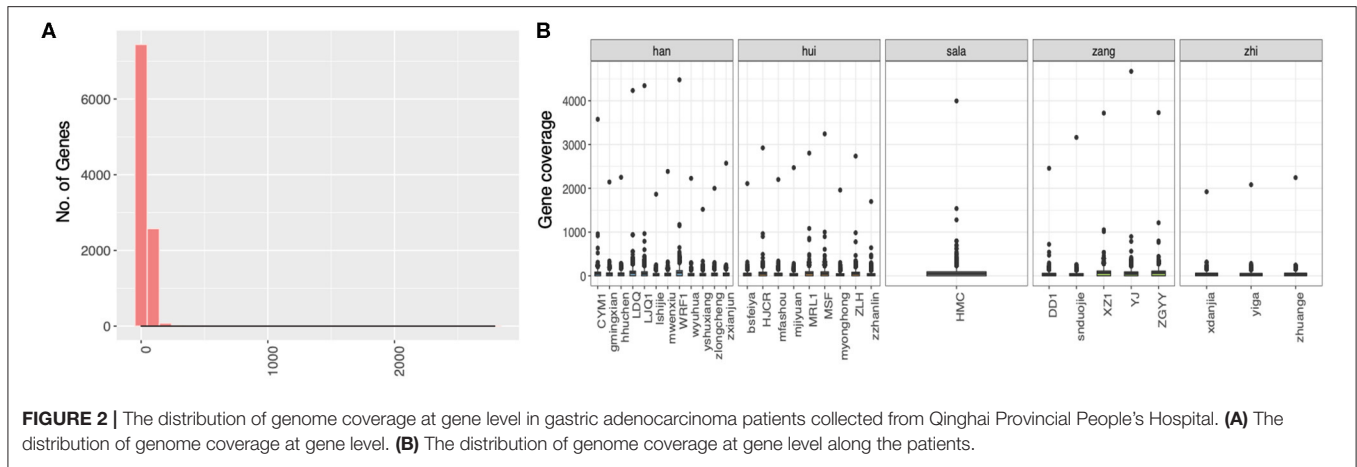


FIGURE 1 | The clinical characteristics for gastric adenocarcinoma patients collected from Qinghai Provincial People's Hospital and TCGA data portal. **(A)** The gender, ethnic group, tumor location, and initial diagnosis age for gastric adenocarcinoma patients collected from Qinghai Provincial People's Hospital. **(B)** The gender, ethnic group, tumor location, and initial diagnosis age for gastric adenocarcinoma patients collected from TCGA data portal.

chromosome 7 for plateau patients. According to the database of COSMIC (31), the world's largest and most comprehensive resource for exploring the somatic mutations in human cancer, chromosome 7 consists of lots of well-known cancer-related somatic mutations, including EGFR, BRAF, CDK6, MET, T1F1, and so on. The copy number variation in chromosome 7 was also related with cancer (32). These results together indicate the important region was detected and the biomarkers for further diagnosis and treatment can be found in chromosome 7. As for SNP alteration type, both western and plateau patients show the majority of C->T and G->A alteration types, and compared with western patients, the variations in each alteration type are bigger in plateau patients than that in western patients (Figures 3B,D).

Using MaxMIF on above somatic mutation data and PPIs from humanNet (27), the first 30 genes with highest MaxMIF

scores were selected as the cancer driver genes (Figure 4). We also performed MaxMIF on western patients, and selected the top 30 genes as the cancer driver genes (Figure S1). Only 5 genes (TP53, RYR2, RYR1, COL12A1, DST) were presented in cancer driver gene list for both plateau and western patients, meaning the molecular differences in plateau patients and western patients. From Figure 4, the two different groups of patients can be found. One is for minor ethnic groups, and the other is Han Chinese and a few of minor ethnic groups of patients, suggesting the difference molecular properties of Han Chinese with minor ethnic groups. In addition, the gene fusion analysis shows in-frame gene fusion of the number 2 exon of KRTAP10-7 and number 1 exon of KRTAP10-6, and number 28 exon of IPO4 and number 21 exon of DNHD1 (Table S2). It also identified the frame-shift gene fusion of the number 1



exon of AKR7A3 and number 11 exon of AKR7L. However, TCGA Research Network only reported in-frame gene fusion of CLDN18 and ARHGAP in western patients (12).

In sum, the unique clinical and molecular characteristics for plateau patients were detected, meaning that there should have some treatment strategies that are specifically designed for them.

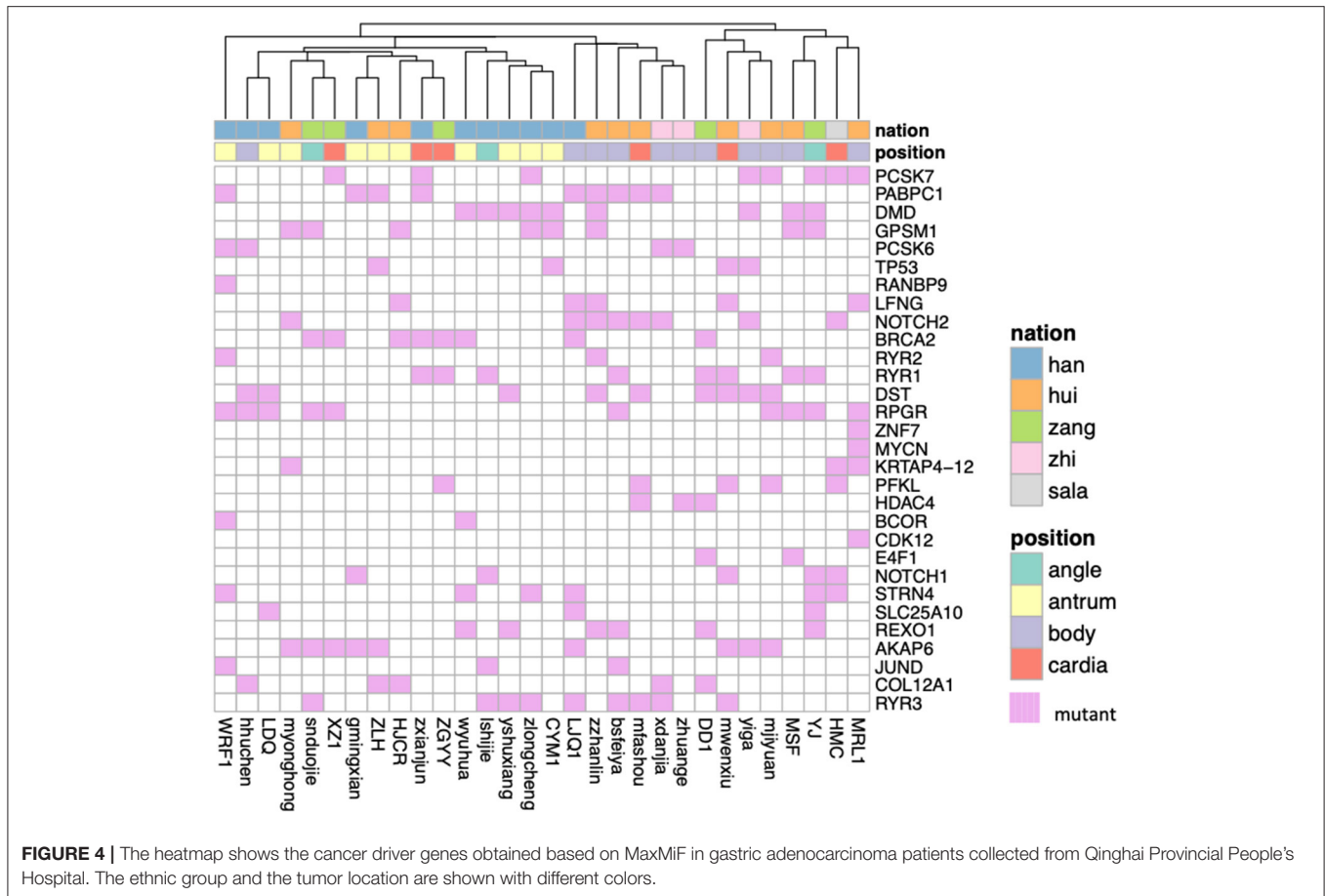


FIGURE 4 | The heatmap shows the cancer driver genes obtained based on MaxMIF in gastric adenocarcinoma patients collected from Qinghai Provincial People's Hospital. The ethnic group and the tumor location are shown with different colors.

Prediction of Effective Clinical Therapies in Plateau Patients

The big challenge in cancer patient treatment is the intratumoral heterogeneity (ITH), a phenomenon observed in a number of aggressive cancers which consist of heterogeneous cell populations with a wide range of morphologies, genotypes and phenotypes. Increasing evidences suggested that ITH plays an important role in drug resistance (33–35). Thus, to detect the effective clinical drugs for these plateau patients that have different clinical and molecular characteristics with western patients, we trained our previously adapted kernel-based learning model (28) on pharmacogenomics generated from patient-derived cancer cells (gliomaScan) that could illustrate the phenomenon of ITH in terms of drug response (13). To evaluate the prediction model, we introduced the clinical drug responses on TCGA cancer patients as the independent testing data. Compared with using cancer genomics generated from cancer cell lines (CTRP) as the training data, trained on cancer genomics generated from patient-derived cancer cells obtained more accurate clinical responses for TCGA patients, no matter using which types of machine-learning algorithms (Figure S2).

Then we applied above prediction model to predict the effective drugs for our plateau patients. The predictions are well-grouped as the types of kinase inhibitors (Figure 5). Not

as the clusters obtained on cancer driver genes, there are no clear classification among Han Chinese and minor ethnic groups of patients, and they will share some target therapies. Among all 30 patients, we found one special minor ethnic group patient (Tibetan people, zang) that was predicted to be treated by Bcr-Abl inhibitor, Bosutinib. Bosutinib is approved to treat chronic myelogenous leukemia (CML) according to NCI reports, and is usually given after other treatments have failed (36). It is not related with gastric cancer yet. Therefore, it might be the good option for that patient if other treatments failed.

DISCUSSION AND CONCLUSION

Qinghai province is in northeast QTP with average 4,000-m height above sea level. This place has its unique climate features, including low temperature (average 21 degrees of centigrade in summer) and low humidity. Han Chinese and lots of minor ethnic groups (such as hui, zang (Tibetan people), zhi, sala, and so on) live here, and they like to eat mutton and beef. In addition, people living here always have alcohol and cigarette problem. The gastric cancer, which are related with dietary and lifestyles, is the most common cancer types in Qinghai province according to the statistical report in Qinghai Provincial People's

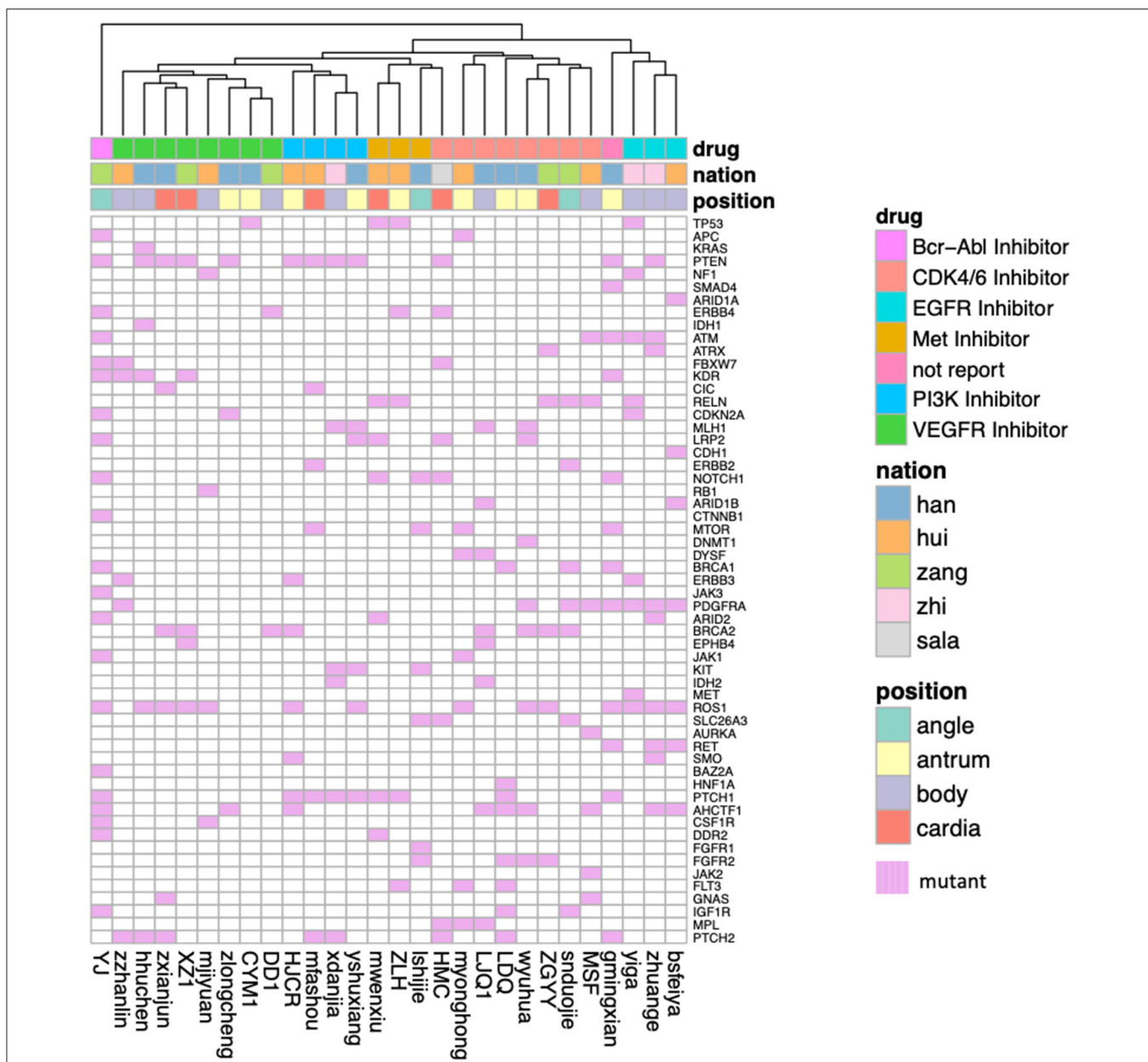


FIGURE 5 | The predicted effective clinical therapy for gastric adenocarcinoma patients collected from Qinghai Provincial People's Hospital. The patients were grouped by the types of kinase inhibitors, and the oncogenes showed here are those somatic mutations in gliomaSCAN. The ethnic group and the tumor location are shown with different colors.

Hospital in 2019. Here, to better understand the molecular mechanism of gastric cancer in QTP, and predict the most effective targeted therapeutics for patients there, we collected the tumor and blood samples from primary gastric adenocarcinoma cancer patients at Qinghai Provincial People's Hospital, and discussed the clinical and molecular characteristics for those patients. By comparing with the western patients collected from TCGA, we found some unique characteristics for our plateau patients, including younger initial diagnosis age, most of tumors located in body of stomach, only share 5 driver genes with TCGA

western patients, have unique gene fusions, et al. Importantly, the cancer driver gene analysis indicates the differences in Han Chinese and minor ethnic groups. For instance, only one of 12 Han Chinese have NOTCH2 mutant, and 7 of 18 minor ethnic people have that gene mutant. It is worthy for further validation by wet-lab *in vivo* and *in vitro* in the future work. In addition, drug response prediction mode based on pharmacogenomics from patient-derived cancer cells suggests the effective targeted therapies, which specifically designed for these plateau patients. The prediction results could be evaluated

by following up tracking reports. In future, we will collect more patients and their follow-up treatment information to validate the predictions here. In addition, the tumor bio-samples were returned to our lab after sequencing. Thus, we will perform the drug-screening assay on these samples to further validate our predicted therapies.

There are around 5,000,000 people living in Qinghai province, and based on the recent statistic reports, only around 1,000,000 people live in Xining. The sample and follow-up treatment information collections are quite challenge here. It took us 2 years to collect these 30 patients. In future, we will collect much more patients with the treatment reports to further show the unique characteristics of plateau patients and discuss their special treatment strategy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Qinghai Provincial People's Hospital. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

1. Wright NA, Poulson R, Stamp G, Van Noorden S, Sarraf C, Elia G, et al. Trefoil peptide gene expression in gastrointestinal epithelial cells in inflammatory bowel disease. *Gastroenterology*. (1993) 104:12–20. doi: 10.1016/0016-5085(93)90830-6
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. (2011) 61:69–90. doi: 10.3322/caac.20107
3. Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol*. (2019) 14:26–38. doi: 10.5114/pg.2018.80001
4. Sitarz R, Skierucha M, Mielko J, Offerhaus GJA, Maciejewski R, Polkowski WP. Gastric cancer: epidemiology, prevention, classification, and treatment. *Cancer Manag Res*. (2018) 10:239–48. doi: 10.2147/CMAR.S149619
5. Rowley DB, Currie BS. Palaeo-altimetry of the late eocene to miocene lunpola basin, central tibet. *Nature*. (2006) 439:677–81. doi: 10.1038/nature04506
6. Favre A, Martin P, Steffen UP, Sonja CJ, Dieter U, Ingo M, et al. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev*. (2015) 90:236–53. doi: 10.1111/brv.12107
7. Xing YW, Richard HR. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc Natl Acad Sci USA*. (2017) 114:E3444–51. doi: 10.1073/pnas.1616063114
8. Spicer RA, Nigel BWH, Mike W, Alexei BH, Guo SX, Paul J, et al. Constant elevation of southern Tibet over the past 15 million years. *Nature*. (2003) 421:622–4. doi: 10.1038/nature01356
9. Fang X, Zhao ZJ, Li JJ, Yan M, Pan B, Song CH, et al. Magnetostratigraphy of the late Cenozoic Laojunmiao anticline in the northern Qilian Mountains and its implications for the northern Tibetan Plateau uplift. *China Earth Sci*. (2005) 48:1040–51. doi: 10.1360/03yd0188
10. Wang Y, Zheng J, Zhang W, Li S, Liu X, Yang X, et al. Cenozoic uplift of the Tibetan Plateau: evidence from the tectonic-sedimentary

AUTHOR CONTRIBUTIONS

GR and YZ conceived the idea and supervised the study. YW designed and implemented the prediction model, and performed data analysis. GR and YZ analyzed the results and drafted the manuscript. SC edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 11671396, 11371365, and 31270270), and a grant from Qinghai Sciences and Technology Department for Basic Research Program (Nos. 2020-ZJ-719 and 2017-ZJ-Y14).

ACKNOWLEDGMENTS

We would like to thank the doctors at Qinghai Provincial People's Hospital for specimen collecting and management. Also, we would like to thank all editors for helping process our paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2020.01033/full#supplementary-material>

- evolution of the western Qaidam Basin. *Geosci Front*. (2012) 3:175–87. doi: 10.1016/j.gsf.2011.11.005
11. Deng T, Ding L. Paleoaltimetry reconstructions of the Tibetan Plateau: progress and contradictions. *Natl Sci Rev*. (2015) 2:417–37. doi: 10.1093/nsr/nwv062
12. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. (2014) 513:202–9. doi: 10.1038/nature13480
13. Lee JK, Liu Z, Sa JK, Shin S, Wang J, Bordyuh M, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat Genet*. (2018) 50:1399–411. doi: 10.1038/s41588-018-0209-6
14. Ding Z, Zu S, Gu J. Evaluating the molecule-based prediction of clinical drug responses in cancer. *Bioinformatics*. (2016) 32:2891–5. doi: 10.1093/bioinformatics/btw344
15. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. (2016) 12:109–16. doi: 10.1038/nchembio.1986
16. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. (2012) 483:603–7. doi: 10.1038/nature11003
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 genome project data processing subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. (2010) 20:1297–303. doi: 10.1101/gr.107524.110

20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010) 38:e164. doi: 10.1093/nar/gkq603
21. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* (2010) 26:841–2. doi: 10.1093/bioinformatics/btq033
22. Ge H, Liu K, Juan T, Fang F, Newman M, Hoek W. FusionMap: detecting fusion genes from next-generation sequencing data at basepair resolution. *Bioinformatics.* (2011) 27:1922–8. doi: 10.1093/bioinformatics/btr310
23. Hou YN, Gao B, Li GJ, Su ZC. MaxMIF: a new method for identifying cancer driver genes through effective data integration. *Adv Sci.* (2018) 5:1800640. doi: 10.1002/adv.201800640
24. Cho A, Shim JE, Kim E, Supek F, Lehner B, Lee I. MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biol.* (2016) 17:129. doi: 10.1186/s13059-016-0989-x
25. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* (2014) 505:495–501. doi: 10.1038/nature12912
26. Lawrence M, S., Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature.* (2013) 499:214–8. doi: 10.1038/nature12213
27. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* (2011) 21:1109–21. doi: 10.1101/gr.118992.110
28. Wang Y, Fang J, Chen S. Inferences of drug responses in cancer cells from cancer genomic features and compound chemical and therapeutic properties. *Sci Rep.* (2016) 6:32679. doi: 10.1038/srep32679
29. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Tist.* (2011) 2:1–27. doi: 10.1145/1961189.1961199
30. Gribskov M, Robinson NL. Use of receiver operating characteristic (roc) analysis to evaluate sequence matching. *Comput Chem.* (1996) 20:25–33. doi: 10.1016/S0097-8485(96)80004-0
31. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* (2019) 47:D941–7. doi: 10.1093/nar/gky1015
32. Kurscheid S, Bady P, Sciuscio D, Samarzija I, Shay T, Vassallo I, et al. Chromosome 7 gain and DNA hypermethylation at the HOXA10 locus are associated with expression of a stem cell related HOX-signature in glioblastoma. *Genome Biol.* (2015) 16:16. doi: 10.1186/s13059-015-0583-7
33. Gerlinger M, Swanton C. How darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer.* (2010) 103:1139–43. doi: 10.1038/sj.bjc.6605912
34. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med.* (2012) 4:127ps10. doi: 10.1126/scitranslmed.3003854
35. Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer.* (2013) 108:479–85. doi: 10.1038/bjc.2012.581
36. Cortes JE, Kantarjian HM, Brümmendorf TH, Kim DW, Turkina AG, Shenm ZX, et al. Safety and efficacy of bosutinib (SKI-606) in chronic phase Philadelphia chromosome-positive chronic myeloid leukemia patients with resistance or intolerance to imatinib. *Blood.* (2011) 118:4567–76. doi: 10.1182/blood-2011-05-355594

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Rong, Zhang, Ma, Chen and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.