

Genomic Landscape and Immunological Profile of Glioblastoma in East Asian Patients

Jean Zhao (✉ jean_zhao@dfci.harvard.edu)

Dana-Farber Cancer Institute <https://orcid.org/0000-0002-4561-5688>

Sheng Zhong

Cancer Hospital of Sun Yat sen University, Dana-Farber Cancer Institute and Harvard Medical School

Bo Wu

The First Hospital of Jilin University

Frank Dubois

Dana-Farber Cancer Institute and Harvard Medical School and Charité - Universitätsmedizin Berlin

<https://orcid.org/0000-0002-7654-6208>

Shanshan Jiang

China Academy of Science

Davy Deng

Dana-Farber Cancer Institute and Harvard Medical School

Hui Li

The First Hospital of Jilin University

Yida Peng

Jilin University

Junliang Ge

The First Hospital of Jilin University

Luwei Liu

Harvard Medical School

Jingyi Liu

Harvard Medical School

Maojin Yao

The First Affiliated Hospital of Guangzhou Medical University

Sean Lawler

Brigham and Women's Hospital

Zhongping Chen

Cancer Hospital of Sun Yat sen University

Tao Jiang

Beijing Neurosurgical Institute <https://orcid.org/0000-0002-7008-6351>

Rameen Beroukhim

Broad institute

Yonggao Mou

Collaborative Innovation Center for Cancer Medicine

Article

Keywords:

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1120759/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

While it is well known that Glioblastoma (GBM) shows profound inter- and intra-tumoral heterogeneity, disparities in molecular features across ancestry groups have been largely overlooked. We collected a large cohort of GBM samples from East Asian patients (EAS-GBM) and performed genomic and transcriptomic analyses of these samples (EAS-GBM, n=443). Further characterization and comparative analysis of the EAS-GBM with the predominantly European-ancestry TCGA GBM dataset (EUR-GBM, n=383) revealed differential genomic and genetic landscape and immunological profile of EAS-GBM from EUR-GBM. EAS-GBM showed an enrichment for NF1, H3F3A, TP53 and ATRX mutations compared to EUR-GBM. Transcriptomic clustering revealed distinct EAS-GBM specific subtypes, namely Proliferative (PL), Neuron-Synaptic (NS), Metabolic (MB) and Immunomodulatory (IM). Notably, the classic subtype of EUR-GBMs with EGFR as its main marker gene was absent in EAS-GBMs, Moreover, the IM subgroup in EAS-GBM with an expression profile that has been previously associated with response to immunotherapy in cancers. Together, our comprehensive characterization revealed the unique genomic and genetic features of EAS-GBMs, providing mechanistic rationale in patient stratification for molecular targeted therapy and drug development in the era of personalized medicine.

Introduction

Glioblastoma multiforme (GBM) is the most common primary malignant tumor in the central nervous system, as well as the leading cause of primary brain cancer-related death. To date, most molecular characterizations of GBM have focused on populations of mostly European ancestry (EUR-GBM), such as The Cancer Genome Atlas (TCGA) program^{1,2}. Therefore, genetic and genomic differences between EAS-GBM and EUR-GBM remain underexplored³. Consequently it remains unclear which factors influence outcomes in EAS-GBM. Here we analyzed specimens from 443 individuals with East Asian ancestry by whole-exome sequencing (WES) and RNA sequencing (RNA-Seq). In this work we will address the following specific questions: What are the differences in genetic drivers between EAS and EUR-GBM? Can we refine the known GBM expression subtypes with the additional data from EAS-GBM? What could be clinically relevant differences in the immune microenvironment and driver mechanisms between these novel molecular subtypes?

Comparison of EAS-GBM genomes to the 383 patients of European descent in the TCGA GBM cohort revealed significant differences in the frequency of well known drivers between EAS and EUR-GBMs and a novel EAS-GBM transcriptomic subgroups. These subgroups include a newly identified immune infiltrated subgroup and show potential clinical utility. Our studies provide comprehensive characterizations of the unique genomic and genetic features of EAS-GBMs, laying a foundation for better understanding ethnicity-dependent tumorigenesis and improved clinical stratification and treatment strategy.

Results

Patient characterization

We recruited 471 adult patients with histologically confirmed GBM from multiple medical centers in China, with the ancestries across East Asian areas, including China, Korea and Japan. Among these, samples from 443 patients passed quality control and were kept for analysis, including 49 de-novo whole-exome sequenced tumors from Sun Yat-Sen University Cancer Center and Jilin University First Hospital for this study and 422 samples were sequenced and re-analyzed from our previous collection ⁴. A subset of 109 tumors from patients over the age of 23 were analyzed with whole exome sequencing (WES) ⁵. 358 GBM patients and 20 normal brain controls were sequenced by RNA-Seq. Among all these participants, 62 patients have both WES and RNA-seq data. The mean coverage for WES data is 127x (range from 89x to 147x), the average number of RNA-seq reads is 45.3 million (range from 27.8 million to 92.3 million).

The EAS-GBMs showed a significant enrichment for IDH mutations (20% IDH^{mut} in EAS vs. 5% in EUR, $q=4\times 10^{-7}$, **Figure 1A**), possibly due to differences in selection criteria. TCGA focused on primary GBM; IDH mutations are more prevalent in secondary GBM ⁶. EAS-GBM patients were significantly younger than the EUR-GBM patients in the TCGA cohort (50 vs. 61, Two-sided Wilcoxon's test, $p<2.4\times 10^{-7}$, **Figure 1B**). This age difference was consistent with previously published studies about GBM in East Asians ^{7,8}. Clinical and molecular data include Primary/Recurrent/Secondary (PRS) type, gender, radiotherapy status, chemotherapy status, IDH mutation status, 1p19q codeletion status, MGMT promoter methylation status and age of onset for a total of all recruited individuals as presented in Table 1, **Figure 1B** and **Extended Table 1**.

Table 1
Characterization Platforms and Data Availability

	EAS GBM	EUR GBM
Data Source	Cases from Tiantan Hospital, Sun Yet-San University Cancer Center and Jilin University First Hospital	Cases from National Cancer Institute, National Human Genome Research Institute
All	443	383
Age, Years		
Median (Min, LQ, UQ, Max)	50 (8,38,58,79)	61 (21,52,69,89)
No. ≤40 years old	115	31
NA	27	8
Glioma Grade		
WHO IV	443	383
GBMs' Prognosis Censor		
Alive	64	72
Dead	328	306
NA	51	5
PRS Type		
Primary	198	/
Recurrent	130	/
Secondary	30	/
NA	85	/
Gender		
Female	165	140
Male	242	241
NA	36	2
Radio Status		
Treated	292	120
Un-treated	91	34

	EAS GBM	EUR GBM
NA	60	229
Chemo Status		
Treated	299	102
Un-treated	87	62
NA	57	219
IDH Mutation Status		
Wildtype	337	356
Mutant	97	21
NA	9	6
1p19q Codeletion Status		
Non-codel	334	/
Codel	20	/
NA	89	/
MGMT promoter Methylation Status		
Methylated	176	17
Un-methylated	138	133
NA	129	233

To assess differences in the driver combinations between East-Asian (EAS) GBM and GBMs of European descent (EUR-GBM), we compared EAS-GBM variants to variant calls for 336 GBMs of European descents downloaded from TCGA. We had RNA seq available from 160 TCGA samples bringing the total size of the EUR-GBM cohort to 383 tumor samples from patients with European ancestry.

Somatic single nucleotide variants and copy number alterations show EAS specific driver combinations

We identified a total of 48,208 somatic single-nucleotide variants (SNVs) and 1,627 insertions/deletions (Indels) in the EAS-GBM cohort, and 71,053 SNVs and 3,203 Indels in the EUR-GBM cohort. Overall, EAS IDH^{wt} GBM showed a significantly higher tumor mutation burden (TMB) than EUR IDH^{wt} (median: 1.19/MB vs. 0.95/MB, Kruskal Wallis test: p=0.02), no statistical significant difference were observed between EAS IDH^{mut} and EUR IDH^{mut} GBM (median: 1.36/MB vs. 0.94/MB, Kruskal Wallis test: p=0.13). (**Extended Material 1-2**). Our analysis of focal somatic copy number aberrations EAS-GBM confirmed

several driver variants known from EUR-GBM including amplifications at 12q14.1(CDK4), 4q12 (PDGFRA), 2p24.3 (MYCN) and deletions at 9p21.3 (CDKN2A). EAS-GBM also showed significant amplifications at 2q11.1 (harboring ANKRD36C) and 6p21.1 (VEGFA), that have not been observed in EUR-GBM (**Extended Figure 1A and Extended Table 2**).

Focusing on significantly recurring driver genes in EAS-GBM (as defined by MutsigCV) we confirmed many of the genes that had been identified as drivers in EUR-GBM (**Figure 1C-D, Extended Figure 1B-D**), including the most frequently mutated driver genes: TP53, ATRX, PTEN, NF1, and EGFR. Our analysis also showed several potentially novel driver genes exclusively observed in EAS GBM (including SOX1 and MEGF9) which will require further validation. When controlling for IDH status, IDH^{mut}-GBM showed no significant differences while the IDH^{wt} EAS-GBM showed a significantly higher rate of H3F3A, NF1, TP53, and ATRX mutations and a depletion of EGFR mutations ($q=0.07$, **Figure 1D, Extended Figure 1B-C**).

TP53 shows a different mutation pattern in EAS- vs. EUR-GBM

In addition to the increased frequency of TP53 mutation in EAS-GBM we also noticed that the location of these SNVs was different. Most of the TP53 variant sites in EUR-GBM scatter evenly in the DNA-binding domain and the oligomerization domain (**Figure 2A-B**), whereas SNVs in variant sites in EAS-GBM concentrated around N-terminal domain residues 1-130 aa, mainly the transactivation domain and the proline-rich domains. Indeed, SNVs were significantly enriched in the transactivation and proline-rich compared to the rest of the protein in EAS-GBM (Fisher test: $p\text{-value}=8.29\times 10^{-4}$ and 6.79×10^{-4} respectively), whereas in EUR GBM the DNA-binding domain showed significant SNV enrichment (Fisher test: $p\text{-value}=1.66\times 10^{-14}$). No significant differences were observed in the oligomerization domain. While the underlying mechanism still needs further elucidation, these data suggest that EAS-GBM is enriched in p53 mutants with different functions than EUR GBM.

Mutational signatures in EAS-GBM show an enrichment for SBS6

We also performed mutational signatures analysis based on the Catalogue of Somatic Mutations in Cancer (COSMIC). Three and four mutational signatures were extracted from EAS and EUR-GBM, respectively (**Figure 2C, Extended Figure 1E**). We found that EAS and EUR-GBM both had 20-30% activity of SBS1 (associated with aging⁹) and SBS3 (associated with deficiency in homologous recombination based DNA repair⁹) (slightly lower in EAS-GBM, $q<0.01$ for SBS1 and $q=0.05$ for SBS3). In contrast, SBS6 was significantly enriched in EAS-GBM ($q<0.01$) (**Figure 2D**). SBS6 showed high activity in hypermutant samples likely due to association with defective DNA mismatch repair (MMR) and microsatellite instability (**Figure 2E**). PARP inhibitors, reported as effective treatment for HRD and MMR tumors, may potentially benefit GBM patients carrying SBS 3 and 6^{10,11}. These data suggest differences in the mutation etiology likely based on differences in environmental exposure and inherent differences in DNA damage response between EAS- and EUR-GBM.

RNA-expression subgroups establish molecular classification for EAS-GBMs

EUR-GBMs (TCGA) were previously classified into three subtypes based on RNA expression: Proneural, Classical and Mesenchymal based on RNA expression clusters^{12,13}, which we confirmed when reanalyzing the TCGA data (**Extended Materials 3-4**). Leveraging our large cohort of RNAseq data from EAS-GBM, we set out to expand this classification.

First, we found that in EAS-GBM IDH^{wt} has a general higher immune response compared to IDH^{mut} GBM (p-value=1.6x10⁻⁶, **Extended Materials 5**) though both IDH^{wt} and IDH^{mut} EAS-GBMs exhibit substantial heterogeneity in immune signature activities. These data suggested that GBM dichotomy solely based on IDH mutation status is insufficient to discriminate intrinsic characteristics of EAS-GBM. Therefore, we conducted a molecular subtype identification for all GBM samples. Unsupervised clustering of 358 EAS-GBM RNAseq samples by non-negative matrix factorization (NMF) consensus clustering divided the cohort into four subtypes (**Extended Figures 2A-B**). These four clusters were also identifiable in t-SNE plots (**Figure 3A**). Clustering was consistent after excluding IDH mutant samples (**Extended Materials 6**), validating the robustness of the methodology we employed and rationality of including both IDH^{wt} and IDH^{mut}. The largest cluster in NMF clustering showed an enrichment for immune related pathways in Metascape pathway analysis. The second largest was enriched in cell cycle, replication and cell division pathways, the third in pathways related to nucleotide metabolism and the smallest cluster was enriched in pathways related to neural synaptic function (**Extended Materials 7, 8**). We therefore named these clusters: Proliferative (PL), Neuron-Synaptic (NS), Metabolic (MB) and Immunomodulatory (IM). PL GBM and IM GBM showed the largest expression distance within EAS-GBM (**Figure 3B**).

Next, we explored whether there are correspondent relationships between EAS and EUR subtypes and subtype taxonomic hierarchies. SubMap results showed that EAS-PL and NS GBM both correspond to EUR Proneural GBM and EAS-IM GBM corresponds to EUR Mesenchymal GBM (**Figures 3C-D**). In contrast, EAS-MB GBM and EUR Classical GBM did not correspond to any subtypes in the other ancestry group, suggesting that they are both unique subtypes for their ancestry, respectively. Marker genes accounting for each GBM subtypes were identified. A substantial number of overlapping marker genes (n=150 and n=69) were found between EUR Proneural and EAS-PL and EAS-NS, respectively (**Extended Figures 2C**). This result supports our previous assertion that both EAS-PL and EAS-NS correspond to EUR Proneural GBM (**Extended Figure 2C**). Similar expression patterns of previously identified EUR proneural GBM marker genes including NF1 and OLIG2 also support this conclusion (**Extended Materials 9**). The clear separations between EAS-PL and EAS-NS, suggests that EUR Proneural GBM could be a mixed group which did not separate due to the limited sample size in the previous study. We identified novel marker genes for PL-GBM (MYB, MYCN, BRIP1, CCND2) and NS-GBM (WIF1, ATP2B3, GRM3, GRIN2A, KIAA1598, FLT3) (**Figure 3E, G, Extended Figure 2D**). Only five marker genes overlap between EAS-MB and EUR classical (**Extended Figure 2C**). EUR Classical GBM marker genes like NOTCH3, TP53, GLI2, SMO, were highly expressed in the EUR cohort (**Extended Materials 9**), but showed low expression in EAS-MB

GBM (**Extended Figure 2D**). This suggests EAS-MB GBM to be a novel subgroup with no correlate in the prior classification. Novel marker genes for EAS-MB GBM include ASPSCR1, NTHL1, MAD1L1, GADD45G, TERT. EAS-IM GBM corresponded to the EUR Mesenchymal group as evident by the large number of overlapping marker genes (n=179) between the two groups (**Extended Figure 2C**). Mesenchymal signature genes like TRADD, CD44, MET and immune related genes containing CTLA4, PDCD1, CD86, CD274, RELB share a similar expression across EAS-IM GBM and EUR Mesenchymal GBM (**Figure 3E, G, Extended Figure 2D, Extended Materials 9**). Our results also show that a series of immune related genes (including REL, TGFB2, COL3A1, COL1A1, SELE, CD274, PDCD1LG2, CCR4, IL21R, IL7R, FAS) are marker genes for IM GBM.

To investigate the underlying biological and clinical application of the new classification system, we studied correlations between GBM molecular subtypes and a variety of molecular and clinical parameters such as PRS type, IDH mutation, 1p19q codeletion, MGMT promoter methylation status, and prognosis (**Figure 3F, Extended Materials 10**). We found that 37/79 of EAS-PL and 19/50 NS GBM have IDH mutation. However, in EAS-IM GBMs IDH mutant cases are dramatically less common (2/92, Fisher test: $p\text{-value}=4.94\times 10^{-5}$). The majority of EAS-PL GBM (54/79) are also MGMT-methylated, whereas the majority of EAS-IM GBM (56/92) are not (Fisher test: $p\text{-value}=1.83\times 10^{-4}$ & 0.04 respectively).

EAS-GBM shows an immune infiltrated subgroup with biomarkers of immune checkpoint therapy response

As immunotherapy continues to play an ever-more important role in the frontiers of cancer therapeutics, we sought to characterize the immunological profiles of each molecular subtype. We found that immune cell infiltration, activation, and inflammation factor level are all higher in EAS-IM GBM than other subtypes. Immune scores, ESTIMATE scores and stromal scores also suggest that immune response is highest in EAS-IM and EUR Mesenchymal GBM (**Figure 4A, Extended Figure 3A-B**). Proneural marker genes, including NF1, NKX2-2, OLIG2, PDGFRA, DLL3 and ASCL1, however show an inverse relationship to the immune infiltration level. Overall our data shows a clear gradient of immune signaling activity with IM-GBM as highest, followed by MB, NF and PL as least immune infiltrated. These data demonstrate different cell fates and signaling axes of the various subtypes.

We also found that HLA genes expressed differently among the four EAS-GBM clusters. EAS-IM GBM generally has the highest HLA gene expression (including HLA-A, HLA-B, HLA-C, HLA-DR). Increased HLA expression could lead to greater antigen presentation and account for the increased immune signatures also observed in EAS-IM GBM. (**Figure 4B, Extended Figure 3C, Wilcoxon tests, all $q<0.06$**). Immune checkpoint blockade therapy target genes, including CTLA4, PDCD1, CD274, and CD86, were highly expressed in EAS-IM GBM relative to other groups ($q<2\times 10^{-2}$, **Extended Materials 11**). The IFN- γ / T-cell-inflamed gene expression profile (GEP) has been shown to be predictive of immune checkpoint blockade (ICB) therapy response in clinical trials in multiple cancers¹⁴⁻¹⁷. EAS-IM GBMs also had the highest IFN- γ /T-cell-inflamed GEP scores compared to the other three subtypes (Wilcoxon test: $p<0.01$, **Figure 4C**), which suggested that EAS-IM GBM patients could potentially benefit more from ICB therapy.

Immune receptor repertoires by EAS subgroup

Accumulated evidence suggests that patients with higher T cell infiltration tend to benefit from ICB treatment¹⁸. To assess the diversity of B and T cells, we employed TRUST¹⁹, which identifies the number of unique clones based on the complementary-determining region 3 (CDR3) of the respective receptor. These data show the TCR CDR3 abundance and diversity to be the highest in EAS-IM GBM, followed by EAS-MB GBM (**Figure 4D, E**). Extra TCR counts and clones could suggest that T cells in IM EAS GBM are more activated compared with the other three subtypes. No significant difference was observed in TCR diversity among EAS-PL, NS and MB GBM. TCR sequence alignment analysis showed a total of five unique TCR CDR3 consensus sequences (C1-C5) in the EAS-GBM micro-environment; such amino acid sequences have been used as templates for engineering CAR-T and TCR-T therapies (EAS-GBM TCR/BCR Repertoire: **Extended Materials 12, Extended Table 3**). $\gamma\delta$ T cells, while only a small percentage of total T cells, are closely associated with anti-tumor immune responses²⁰. Recent studies reported that V γ 9V δ 2, a major subset (65-95%) of $\gamma\delta$ T cells are able to recognize and kill cancer cells through a TCR-dependent manner²¹. Among EAS-GBM, the NS subgroup has the lowest $\gamma\delta$ T cell infiltration whereas IM tumors have the most ($P < 0.05$) (**Extended Materials 13**).

We next investigated BCR repertoires (including heavy chains: IgG, IgA, IgE, IgM, IgD) in the EAS-GBM tumor microenvironment. Normalized BCR CDR3 counts are also significantly higher in the MB and IM GBM relative to the other subgroups (**Figure 4F**, $P < 0.05$). In contrast to TCR diversity, BCR diversity was significantly lower in IM GBM compared to the other groups (**Figure 4G**). Such restricted clonal diversity but increased BCR abundance could imply that specific B cells expanded as part of an adaptive immune response in these GBMs. Among immunoglobulin subtypes (**Extended Materials 14**), the results revealed that IgA1, IgG1, IgG3, IgG4, IgM ratios are remarkably higher in EAS-IM GBM ($P < 0.05$). No obvious difference in IgE and IgG2 fraction was observed in four clusters. These data show that the IM GBMs are characterized by a higher abundance of polyclonal T-cells compared to the other groups. They also show more B-cells of several classes with reduced diversity of the variable regions of their B-cell receptors, indicating possible expansion of tumor specific clones.

Predicting overall survival in EAS-GBM patients

Previous studies have shown that HLA serotype and haplotype has diagnostic value and clinical application in melanoma and non-small cell lung cancers²². We conducted a similar prognostic analysis of HLA serotypes and haplotypes in our EAS-GBM cohort. According to Kaplan-Meier and multivariate survival analysis, loss of heterozygosity (LOH) in HLA supertype (serotype) did not have an impact upon GBM patients' prognosis (**Extended Materials 15, Extended Table 4**).

GBM patients with the HLA supertypes HLA-A*68, HLA-A*31, HLA-B*08, HLA-B*50 or HLA-B*56 (genotype: HLA-A*31:01, HLA-B*50:01, HLA-B*56:01) had a worse prognosis (**Extended Materials 16**). In contrast, patients with HLA-B*54 had better survival (**Extended Materials 16**, log rank test: $p < 0.05$). This was specifically driven by the HLA-B*54:01 genotype (which was only HLA-B54 present in EAS-GBM)

(Extended Figure 3E). This significance is also confirmed by the multivariate Cox regression model by integrating other clinical variates. It is tempting to speculate that HLA-B*54:01 could allow for the ideal presentation of a yet to be identified GBM specific antigen while HLA-A*31, HLA-B*50 or HLA-B*56 completely lack that capacity **(Extended Materials 17)**.

Our analysis revealed that age has significant correlation with multiple factors, including stromal and immune score, HLA expression level, MSI and TMB. Stromal score, immune score, HLA expression level and indices of genome instability (GII, Ploidy, GII Deletion, GII Amplification) all correlated with each other **(Extended Materials 18)**. We therefore decided to carry out log rank survival analysis and multivariate survival analysis **(Extended Figure 3E-3J, Extended Materials 19)**. A multivariate cox analysis found higher ploidy, GII (Genome Instability Index), MATH (Mutant-Allele Tumor Heterogeneity) scores and H3F3A and PAN3 somatic mutations to be associated with unfavorable prognosis in EAS-GBM patients. We also found that the NS subtype independently predicted a better prognosis compared to the other three subtypes **(Extended Figure 3D, Extended Materials 20)** (risk: 0.850, 95% confidence interval: 0.748-0.966, $p < 0.013$). We found no correlation with survival for any of the immune-related indicators (normalized TCR CDR3 counts, TCR clonal diversity and $\gamma\delta$ T cell fractions, normalized BCR CDR3 counts, BCR clone diversity, heavy/light chain subtype) **(Extended Materials 21-22)**.

To allow for rapid assessment of EAS-GBM survival probability in the clinic, we established a nomogram scale plot to calculate 1-, 3- and 5-year-survival probabilities **(Extended Figure 4A-B)**. Receiver operating characteristic (ROC) curves of the nomogram were above 0.7 for both 1- and 5-year survival prediction **(Extended Figure 4C)**. We also trained a multiple Cox regression-based machine learning model to predict EAS-GBM patient prognosis based on multi-omics data, including clinical information, driver gene mutation status, genome stability index, immune score, and gene expression profile **(Extended Figure 4D)**. Patients with a high median risk score were correlated with worse OS as expected **(Extended Figure 4E)**. AUC values also corroborated the high prediction efficiency of our multiple Cox regression model **(Extended Figure 4F)**. The C-index, short for concordance index, is a metric to evaluate the predicted accuracy of our multiple Cox regression model. C-index results indicated that machine learning model integrated multi-omics data (including clinical features, genome index and somatic mutation status) have a significantly better performance than models constructed by any single features **(Extended Figure 4G)**.

Treatment opportunity and efficacy in EAS-GBMs

We also evaluated treatment efficacy of different modalities in each GBM subtype using multivariate Cox models including chemotherapy, radiotherapy, age, gender, disease (PRS) type, IDH mutation status, 1p19q codeletion status, and MGMT promoter methylation status **(Figure 5, Extended Materials 23)**. All EAS-GBM subtypes except NS GBM showed improved outcomes after receiving chemotherapy (all $p < 0.05$ after controlling for the other factors). Meanwhile, EAS-MB GBM appeared to benefit from radiotherapy significantly ($p = 0.001$). Notably, EAS-NS GBM is not sensitive to either chemotherapy or radiotherapy. These data suggest that the EAS-GBM expression subtypes could influence clinical management.

It was reported that tumor cells with homologous recombination defects (HRD) are more sensitive to platinum drugs or PARP inhibitors, such as Olaparib^{23,24}. Our results reveal that only about 5.1% of patients (EAS cohort: 3/109, 2.8%; EUR cohort: 20/336, 6.0%. Fisher test: p-value = 0.22) are predicted to potentially benefit from PARP inhibitors (**Extended Materials 24**) based on somatic mutation calls. About 7.1% (EAS cohort: 10/109, 9.2%; EUR cohort: 22/336, 6.5%. Fisher test: p-value = 0.39) patients show MMR SNVs (**Extended Materials 24**). However, higher TMB patients caused by MMR were reported to not respond to ICB inhibitors²⁵. These results also suggest EAS-GBM molecular subtype rather than TMB could be an ICB response indicator.

Discussion

Many tumor types in TCGA including GBM suffer from the underrepresentation of samples from diverse ancestries leading to barriers to the generalizability and interpretation of research and clinical trial data. Leveraging a large cohort of 443 GBM samples of East Asian descent, we showed genetic and transcriptomic differences between the TCGA European and East Asians GBM. Aside from different frequencies of known drivers, EAS-GBM also showed differences in SNV loci distributions of TP53 variants. In EAS-GBM they are mostly concentrated around the N-terminal transactivation domain while in EUR-GBM they are primarily located in the DNA-binding domain. The p53 N-terminus is responsible for its interaction with MDM2, leading to its degradation^{26,27}, but underlying mechanisms accounting for this phenomenon in EAS-GBM still need further clarification.

We established a novel four-molecular subtype system for EAS-GBM (namely, Proliferative, Neuro-Synaptic, Metabolic and Immunomodulatory) based on RNAseq data from 358 EAS GBMs using NMF classification. Verhaak et al proposed a widely known four-class classifications criteria (containing Proneural, Neural, Classical and Mesenchymal) based on transcriptome data by employing the nearest centroid-based classifier¹². Notably the classical subtype in EUR-GBM was largely absent in EAS-GBM. Interestingly, through driver gene analyses, we found that this absent EUR group is driven by EGFR activating mutations, whose SNVs were found to be depleted in the EAS cohort, indicating a less prominent role for EGFR signaling in EAS GBM than in EUR GBM. Furthermore, EAS-GBM was also characterized by a subgroup that showed immune infiltrate expression signatures. This subgroup contained a third of all EAS-GBM patients and showed expression signatures that have been associated with response to immunotherapy in other cancer types^{14,16}. This could suggest that a subset of EAS-GBM have a higher chance of responding to immunotherapy. We also found that our novel four class EAS GBM classification matches well with previous hierarchical clustering based on GBM scRNA data²⁸. PL, NS, MB and IM are aligned with GBM patients enriched in cell cycle, oligo, hypoxia and immune signaling pathway.

The prior transcription based classifications so far have not made any impact on clinical management. Analyses of the clinical outcomes in the EAS cohort showed that the different expression subgroups respond differently to the current first line standard of care: chemotherapy and radiation. While NS EAS-

GBM showed no benefit from either therapy, IM and PL GBM showed a benefit from chemotherapy only and MB EAS GBM responded to both chemotherapy and radiation. These findings highlight a potential direct clinical relevance of the expression subgroups in EAS-GBM.

Overall, we believe the EAS-GBM classification proposed in this manuscript dissected out some of its unique genetic and transcriptomic differences, both within itself and in relation to EUR-GBM. Additions from multi-omics modalities could further refine this classification, and add markers that can predict differential responses to treatments, particularly those that are disparate in different ancestry. This work could provide guidance ranging from research design, clinical trial recruitment to daily clinical management of EAS-GBM.

Methods

Patient recruiting and characterization

Glioblastoma tissues, corresponding genomic data and patients' clinical information (diagnosis, gender, age, WHO grade and overall survival (OS)) were obtained from 471 GBM patients (GBMs). 422 GBM patients were recruited from Beijing Tiantan Hospital while 49 GBM patients recruited from Sun Yat-Sen University Cancer Center and Jilin University First Hospital. Among all these participants, 109 patients' tumors were performed with whole exon sequencing (WES) and 358 patients and 20 normal brain controls were sequenced by RNA-Seq. After quality control, 463 samples thus remained for the following analysis procedure, 62 patients have both WES and RNA-seq data. All experiments were performed in accordance with the guidelines of the institution ethics committee (AAALAC standard), and were conducted in accordance with principles expressed in the Declaration of Helsinki. This research was approved by ethic committees of the participating institutes. Besides, informed consent was obtained before any performance.

According to the World Health Organization (WHO) 2016 and 2021 classification of central nervous system tumors, patients in this study were diagnosed by three independent experienced pathologists based on histological characteristics combined with molecular characteristics^{29,30}. In this study, glioblastoma were further divided into PRS types, which refers to primary GBM (pGBM), secondary GBM (sGBM) and recurrent glioblastoma (rGBM). pGBM is primary GBM without any clinical or histologic evidence related to less malignant lesions. sGBM in this study refers to glioblastoma that developed from low grade glioma (LGG) by definition, usually accompanied with IDH mutation. In addition, rGBM refers to a patient whose GBM has been diagnosed to relapse again. Considering the impact of IDH mutation status on the prognosis of GBMs, whole exome sequencing (WES) and RNA sequencing (RNA Seq) were performed on 471 individuals of East Asian descent, which were grouped by EAS IDH1wt and IDH1 mutant groups. IDH1 mutation status in this study was detected by Sanger sequencing independently, 1p19q co-deletion status was checked by Fluorescence in situ hybridization (FISH), MGMT promoter methylation status was determined following previous protocol³¹.

Sample preparation and sequencing

Glioblastoma tissues were snap-frozen in liquid nitrogen immediately after surgical resection and preserved in liquid nitrogen. First, we extracted genomic DNA from the tumor and matched blood samples, and confirmed its high integrity by 1% agarose gel electrophoresis. DNA was then fragmented for quality control, and paired libraries were prepared. For whole exome sequencing, Agilent SureSelect kit v5.4 was used for target capture. Sequencing was performed on the Illumina HiSeq 4000 platform using paired-end sequencing strategy. Total RNA was extracted from each tumor sample, followed by removing tRNA and rRNA. After quality control and quantification, mRNA was reverse transcribed into cDNA, then followed by library preparation and sequencing.

Somatic variant identification

Short sequence reads were mapped to the human reference genome GRCh37 using BWA-MEM (v.0.7.17) with default parameters. Following GATK best practice, PCR duplicates were first removed and subsequently realigned and recalibrated (available at <https://github.com/gis-rpd/pipelines>, GATK v.4.1.9.0). Similar steps, including marking duplication, realignment and recalibration, were performed for BAM files. SNVs were identified using MuTect (v.1.1.7). Tumor samples were then used to call somatic mutations against the paired normal. In addition, based on the validation results, we further filtered out false positive mutations and retained only those supported by at least five mutation reads with variant allele frequency (VAF) > 0.05. Somatic insertions and deletions (Indels) were double checked using Strelka (v.1.0.15) with default parameters. All qualified somatic variants were annotated using Oncotator (v.1.9.9.0).

Driver genes identification

To identify driver mutations in EAS and EUR cohorts, the MutSigCV (v.1.41) and Maftools R package were used to infer significantly mutant driver genes ($q < 0.1$ in any caller, and non-silent mutations $n \geq 5$). We annotated GBM drivers with a combination of driver lists: (1) Significantly mutated genes in GBM proposed by previous publications; (2) GBM associated proto-oncogene and tumor suppressor genes in Genomic Data Commons (GDC) database³²; (3) New drivers identified from the EAS cohort in this study.

Mutational signatures deconvolution and identification

To uncover etiological factors that drive GBM oncogenesis and pathogenesis, Maftools and YAPSA R package was used for de novo discovery of mutational signatures in EAS-GBM cohort. The MutationalPatterns R package was then employed to characterize and visualize results from de novo mutational signature inference. The same protocol was performed on the TCGA cohort. Bases mutational patterns in each category were then assigned to single base substitutions (SBS) features as previously reported in Catalogue of Somatic Mutations in Cancer (COSMIC), proportions of each signature were then estimated to investigate its contribution to GBM formation.

Copy number variation (CNV) of EAS and EUR-GBMs

Sequenza (v.3.0.0) was firstly employed to estimate copy number profile (including allele-specific copy number), tumor purity and ploidy for each patient (18). Default settings were used following Sequenza documentation. GISTIC (v.2.0.23) was used to identify significantly amplified and deleted regions. Output segmentation files generated from Sequenza were then used as input for GISTIC (19). GISTIC parameters were set as default according to the manual. Chromosome arms were labeled as 'altered' in each cohort if GISTIC $q < 0.1$. GII was calculated as the percentage of a tumor genome showing a copy number deviated from the median of the genome (all CNVs are integers as inferred by Sequenza). GII deletion was defined as the percentage of a tumor exhibiting significant decreased copy number. Then, microsatellite instability (MSI) indexes were evaluated by MSISensor via inputting alignment bam files.

Transcriptome clustering using non-negative matrix factorization

Raw RNA-seq reads from 388 tumors and 20 normal samples of EAS cohort and 160 tumor samples of EUR cohort were aligned to the reference genome (UCSC hg19 with annotations from GENCODE v.37) using STAR (v.2.7.3a). 30 RNA-Seq samples were removed from the cohort since PCA results detected that potential RNA degradation occurred in these samples. Gene expression profile was subsequently quantified using RSEM (v.1.3.3). Expression counts were estimated by RSEM, TPM were subsequently calculated by dividing the read counts by the length of each gene in kilobases. For clustering of tumor samples, the top 3000 most variable coding genes (based on median absolute deviation) were selected from tumor transcriptome data. To obtain robust clusters consensus, an unsupervised clustering method named non-negative matrix factorization (NMF) was used (NMF R-package v. 0.23.0). Optimal rank parameters were first determined using 50 runs of random starting points with default settings, followed by 300 runs with optimal ranks to acquire the final NMF clustering solutions. Silhouette analysis was performed to assess the consistency of proposed clustering solutions. Clustering significance was evaluated using SigClust, and all class boundaries were proved to be statistically significant.

Principal component analysis and clustering

Principal Component Analysis (PCA) was first employed to detect RNA decomposition of these samples. Next, PCA were performed again to check correlation between proposed molecular subtypes and batches based on top 3000 variable genes used in NMF clustering, PCA was realized using R built-in function `prcomp`. Sample points were colored respectively based on their NMF clusters and batches, while shapes were annotated by PRS type. Besides, t-SNE clustering was performed to check grouping efficiency by applying the t-SNE R package. Phylogenetic tree of the EAS cohort was analyzed to plot the genetic distance among different molecular subtypes using the `monocle` R package (v.0.2.3).

Molecular subgroup mapping across EAS and EUR Cohort

To investigate correlation among molecular subtypes identified from EAS and EUR cohorts, an unsupervised subclass mapping method SubMap (<https://www.genepattern.org/>) was used to identify

correspondence or commonality of subtypes from the two cohorts. SubMap analysis was performed on the GenePattern online platform using default settings with a random seed of 1. The intersection of genes selected for NMF clustering (3,000 most variable genes in each cohort) were input to perform analysis. Significant correspondences were identified with a cutoff of Bonferroni adjusted $P < 0.01$.

Marker genes selection and pathway enrichment analysis

To identify marker genes for each GBM subtypes, Comparative Marker Gene Selection algorithms on the GenePattern online platform were performed with default settings and a threshold score > 0.5 . Top 500 differential expressed genes (DEGs) were selected as marker genes for each subtype, DESeq2 (3.12) in R were then conducted to validate results. To further compare similarity across EAS and EUR subtypes, VENN plots were performed to check the overlapped marker genes on Venn webtools (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

To uncover Gene Ontology (GO) and pathways (Kyoto Encyclopedia of Genes and Genomes, KEGG) enrichment of each GBM subtypes, marker genes of each subtype were respectively uploaded to online tools: Metascape (online websites with gene annotation, visualization: <https://metascape.org/gp/index.html>). Using the fgsea R package (v.3.12), GSEA was performed with assistance of hallmark gene set from MSigDB (v.7.4). Significant enriched gene sets were filtered based on a cutoff of $q < 0.01$.

Immune profile analysis of RNA subtypes

In decomposing immune components of tumor samples, enrichment levels of the 29 immune signatures were quantized using single sample gene set enrichment analysis (ssGSEA v.10.0.2) score, resulting in higher immune scores denoting a greater immune component in a tumor³³. ESTIMATE R package (v.9.1.0) along with CIBERSORTX (<https://cibersort.stanford.edu/>), MCP-Counter (v.1.1) R package as well as online tools TIMER (<http://cistrome.org/TIMER/>) were then employed to check data consistency using default settings. To predict Immune Checkpoint Blockade (ICB) therapy responses, an 18-gene T cell inflamed gene expression profile (GEP) was calculated, it refers to a weighted sum of these 18 genes after normalization with 11 housekeeping genes following previously published methods¹⁴.

HLA supertype and haplotype identification

Human leukocyte antigen (HLA) molecules are highly polymorphic. They contain various supertypes (also known as serotypes) and haplotypes even within the same race, and across different races. Different HLA supertypes and haplotypes have shown a significant disparity in antigen presenting ability and peptide ligands affinity³⁴. Expression of HLA genes among GBM subtypes were obtained and compared from transcriptome TPM value. HLA supertypes and haplotypes were identified using two independent software tools, arcasHLA (v.0.2.5) and seq2HLA with default parameters^{35,36}. Only haplotypes and supertypes confirmed by the two softwares were retained for future analysis. HLA loss of heterozygous (LOH) was considered as an individual that only one allele gene was detected on his or her allele loci.

Association between HLA status and clinical features were then investigated by univariate and multivariate analysis.

Detection and analysis of TCR, BCR CDR3 sequences from GBMs RNA-Seq data

TRUST (v.3.0.1) was applied (<https://bitbucket.org/liulab/trust>) to extract TCR/BCR sequence from 345 GBMs RNA-seq data³³. CDR3 sequences with significantly fewer reads (less than 3) were excluded in this analysis. The number of total sequencing reads was obtained from each bam file using Samtools, each variable (V), joining (J), or constant (C) region of CDR3 sequences were annotated with the help of TRUST. In order to compare the richness of TCR/BCR among different subtypes, we normalized CDR3s counts by total reads and tumor purity in each sample. Clonotype diversity of T/B cells was estimated by TCR/BCR CDR3s per kilo TCR/BCR reads (CPK) in each sample as previously described³⁷. Shannon entropy calculated from the frequencies of TCR β chain and IgH CDR3 sequences is another index evaluating diversity of TCR/BCR, and such an index represents the adaptive ability of patients' immune systems. $\gamma\delta$ T cell fraction, which plays a pivotal role in tumor cytotoxic immunity³⁸, was estimated by the total number of γ or δ CDR3s divided by the total number of TCR CDR3s in each sample³⁹. In identifying B cell lineage clusters in each sample, immunoglobulin heavy/light chain subtypes were aligned by the constant (C) region of IgH CDR3 sequences or Ig light chain; the fraction of each subtype was calculated by dividing heavy/light chain total reads in each sample. TCR, BCR and tumor clones

Shannon Index (SI) are calculated with the formula: $SI = \sum_i^{Clone\ Numbers} -P_i * \ln P_i$, where P_i is the proportion of clone i ⁴⁰.

Survival analysis and correlation of multilayer features

Univariate and multivariate Cox analysis were performed to check correlations between prognosis and features (clinical features, including Chemo-/ Radiotherapy Status, Molecular Subtypes and PRS Types; genomic features, including somatic mutation of genes, Ploidy and GII). Survival outcomes of different groups were estimated by the Kaplan-Meier method and log-rank tests were used to calculate P values. To uncover the correlation of multilayer features in EAS cohort, a list of multiple layers features was curated, including baseline information (including age, sex and others), treatment regimen (chemotherapy status, radiotherapy status), genome indices (Ploidy, GII, GII deletion), somatic mutation of diver genes, molecular subtype. Pairwise correlations of all features were surveyed using the GGally R package (v.1.5.0) and corrgram R package (v.1.13.). MATH (Mutant-Allele Tumor Heterogeneity) scores were calculated to assess intratumor heterogeneity among GBM samples, it's calculated as:

$MATH_i = \frac{MAD(VAF_i)}{Median(VAF_i)} \times 100$, where VAF_i is a vector of the VAFs of all mutations from sample i and MAD denotes median absolute deviation⁴¹.

Construction and validation of prognostic signature and nomogram

Based on patients' clinical, genomic features and molecular subtypes, an innovative nomogram specific to EAS-GBM was first developed by rms r package (v.6.2-0.). Calibration charts were performed to determine 1-, 3- and 5- year OS prediction accuracy of our nomogram. Sensitivity and specificity of nomogram were assessed using survivalROC R package. Then, a machine learning predicting model was constructed. Cox regression coefficients were firstly utilized to estimate risk scores for each patient. Patients were then divided into two groups (low risk and high risk) according to their risk score: The boundary value was set by the gaurvminer R package (v.0.4.9). Subsequently, survival outcomes of the two groups were evaluated by Kaplan-Meier prognosis plot and a log-rank test was used to calculate P values. Receiver operating characteristic (ROC) curve and area under curve (AUC) value were calculated via the survivalROC R package (version 1.0.1). Consistency indexes (C index) were then calculated with randomly performing 10-fold cross-validation 100 times to assess model prediction efficiency in each input setting.

Somatic mutation pathway annotation

Somatic mutation pathway enrichment was annotated and summarized by Maftools, then the results were visualized by the online tool Pathway Mapper (<http://www.pathwaymapper.org/>). Homologous recombination deficiency (HRD) refers to an incorrect alignment of chromosomes during meiosis, which is strongly linked to cancer formation. We selected a panel of genes, including BRCA1/2, CDK12, ATM, MLH1/3, MSH2/3/6, PMS1/2 and PALB2 mutation to evaluate patients' HRD status. DNA mismatch repair (MMR) pathways play key roles in proofing of errors during DNA replication, MMR associated genes including MLH1, MLH3, MSH2, MSH6, PMS1 and PMS2 were investigated to determine MMR status.

Statistics

The Wilcoxon rank-sum test and Kruskal-Wallis test were used to check the difference in significance of continuous variables among different groups. Spearman's rank correlation was employed to check correlation between continuous variables. Chi square or fisher's exact tests were applied to test for enrichment of categorical variables among different groups. Multiple testing was accounted for by using false discovery rate q-values unless otherwise indicated⁴². P<0.05 was set to be statistically significant across all the tests above.

References

1. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068 (2008).
2. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477 (2013).
3. Carrot-Zhang, J. *et al.* Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell* 37, 639–654.e6 (2020).

4. Zhao, Z. *et al.* Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. *Genomics Proteomics Bioinformatics* 19, 1–12 (2021).
5. Paugh, B. S. *et al.* Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. *J. Clin. Oncol.* 28, 3061–3068 (2010).
6. Han, S. *et al.* IDH mutation in glioma: molecular mechanisms and potential therapeutic targets. *Br. J. Cancer* 122, 1580–1589 (2020).
7. Zhang, G.-B. *et al.* Differential molecular genetic analysis in glioblastoma multiforme of long- and short-term survivors: a clinical study in Chinese patients. *J. Neurooncol.* 113, 251–258 (2013).
8. Li, Q. *et al.* The analysis of EGFR variants in Chinese adult glioblastoma patients. *J. Clin. Oncol.* 39, e14020–e14020 (2021).
9. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101 (2020).
10. Miller, R. E. *et al.* ESMO recommendations on predictive biomarker testing for homologous recombination deficiency and PARP inhibitor benefit in ovarian cancer. *Ann. Oncol.* 31, 1606–1622 (2020).
11. Higuchi, F. *et al.* Restoration of Temozolomide Sensitivity by PARP Inhibitors in Mismatch Repair Deficient Glioblastoma is Independent of Base Excision Repair. *Clin. Cancer Res.* 26, 1690–1699 (2020).
12. Verhaak, R. G. W. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110 (2010).
13. Wang, Q. *et al.* Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell* 33, 152 (2018).
14. Ayers, M. *et al.* IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* 127, 2930–2940 (2017).
15. Cristescu, R. *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* (2018).
16. Chalabi, M. *et al.* Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat. Med.* 26, 566–576 (2020).
17. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* 19, 133–150 (2019).
18. Jiang, P. *et al.* Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24, 1550–1558 (2018).
19. Song, L. *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* 18, 627–630 (2021).
20. Yazdanifar, M., Barbarito, G., Bertaina, A. & Airoidi, I. $\gamma\delta$ T Cells: The Ideal Tool for Cancer Immunotherapy. *Cells* 9, (2020).

21. Gertner-Dardenne, J. *et al.* Human V γ 9V δ 2 T cells specifically recognize and kill acute myeloid leukemic blasts. *J. Immunol.* 188, 4701–4708 (2012).
22. Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 359, 582–587 (2018).
23. Telli, M. L. *et al.* Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. *Clin. Cancer Res.* 22, 3764–3773 (2016).
24. Cerrato, A., Morra, F. & Celetti, A. Use of poly ADP-ribose polymerase [PARP] inhibitors in cancer cells bearing DDR defects: the rationale for their inclusion in the clinic. *J. Exp. Clin. Cancer Res.* 35, 179 (2016).
25. Touat, M. *et al.* Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* 580, 517–523 (2020).
26. He, F. *et al.* Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* 116, 8859–8868 (2019).
27. May, P. & May, E. Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene* 18, 7621–7636 (1999).
28. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401 (2014).
29. Gritsch, S., Batchelor, T. T. & Gonzalez Castro, L. N. Diagnostic, therapeutic, and prognostic implications of the 2021 World Health Organization classification of tumors of the central nervous system. *Cancer* (2021) doi:10.1002/cncr.33918.
30. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* 131, 803–820 (2016).
31. Wang, L. *et al.* Comparative assessment of three methods to analyze MGMT methylation status in a series of 350 gliomas and gangliogliomas. *Pathol. Res. Pract.* 213, 1489–1493 (2017).
32. Zhang, Z. *et al.* Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat. Commun.* 12, 1226 (2021).
33. Zhang, J. *et al.* Immune receptor repertoires in pediatric and adult acute myeloid leukemia. *Genome Med.* 11, 73 (2019).
34. Wang, M. & Claesson, M. H. Classification of human leukocyte antigen (HLA) supertypes. *Methods Mol. Biol.* 1184, 309–317 (2014).
35. Orenbuch, R. *et al.* arcashLA: high-resolution HLA typing from RNAseq. *Bioinformatics* 36, 33–40 (2020).
36. Boegel, S. *et al.* HLA typing from RNA-Seq sequence reads. *Genome Med.* 4, 102 (2012).
37. Li, B. *et al.* Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat. Genet.* 48, 725–732 (2016).

38. Latha, T. S. *et al.* $\gamma\delta$ T Cell-Mediated Immune Responses in Disease and Therapy. *Front. Immunol.* 5, 571 (2014).
39. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
40. Merlo, L. M. F. *et al.* A comprehensive survey of clonal diversity measures in Barrett’s esophagus as biomarkers of progression to esophageal adenocarcinoma. *Cancer Prev. Res.* 3, 1388–1397 (2010).
41. Chen, J. *et al.* Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* 52, 177–186 (2020).
42. Storey, J. D. False Discovery Rate. *International Encyclopedia of Statistical Science* 504–508 (2011) doi:10.1007/978-3-642-04898-2_248.

Figures

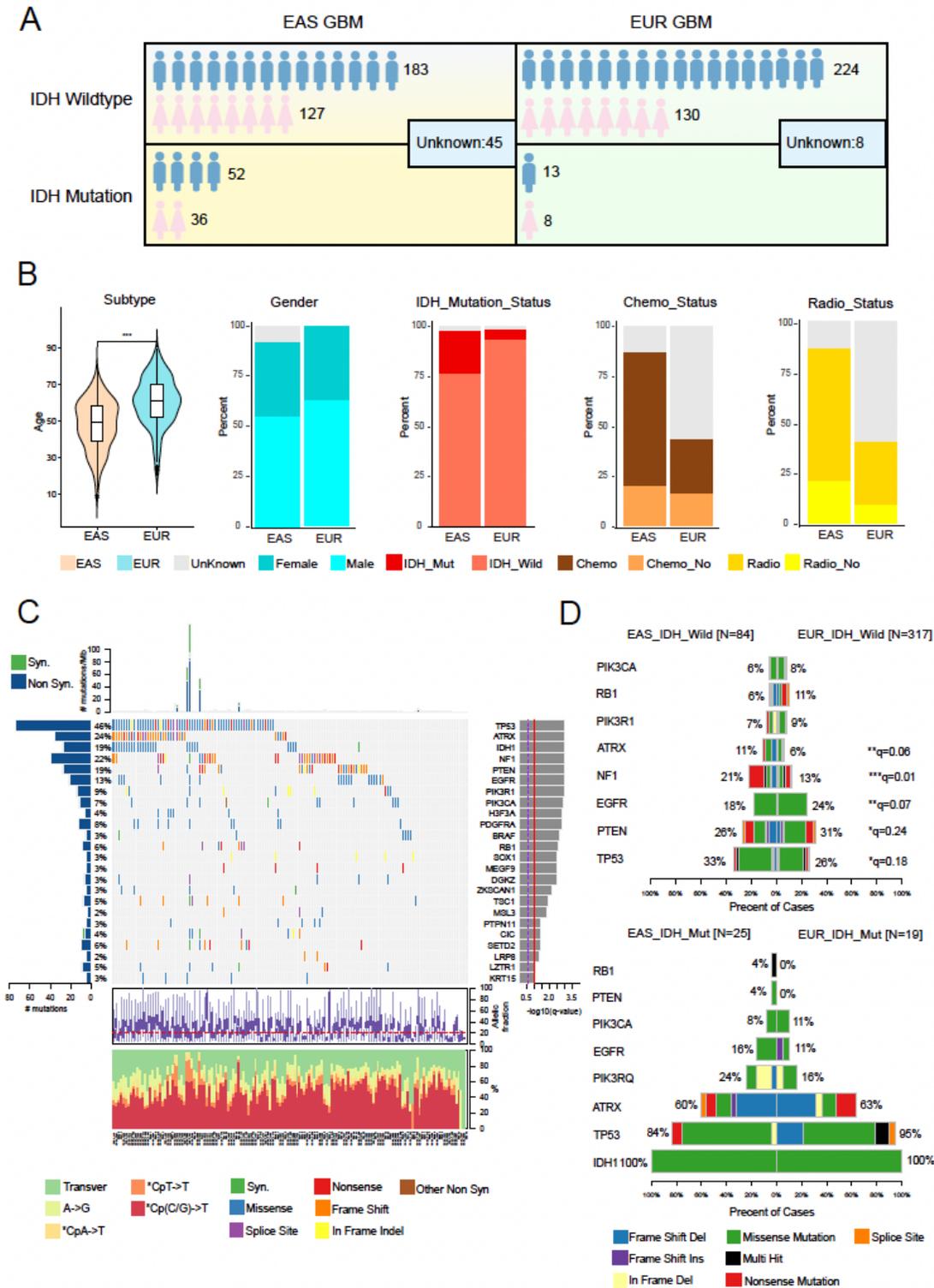


Figure 1

(A) Cohort composition of the EAS and EUR datasets in this study. (B) Differences between EAS and EUR GBM in age, gender, therapy status and IDH mutation status. (***) p -value < 0.01). (C) Genes with significantly recurring mutations in EAS-GBM ($q < 0.1$). The top row indicates the mutational load of each patient. The following rows show the mutation type of each gene in all patients and the significance of each mutation is displayed in the right column. Mutation rates are displayed in the left column. The

bottom row showed the ratio of tumor reads versus total number of reads per sample and mutation spectrum of bases substitutions. (D) Comparison of the prevalence of significantly recurrent mutations in cancer genes between the EAS and EUR-GBMs for IDHwt (top) and IDHmut (bottom).

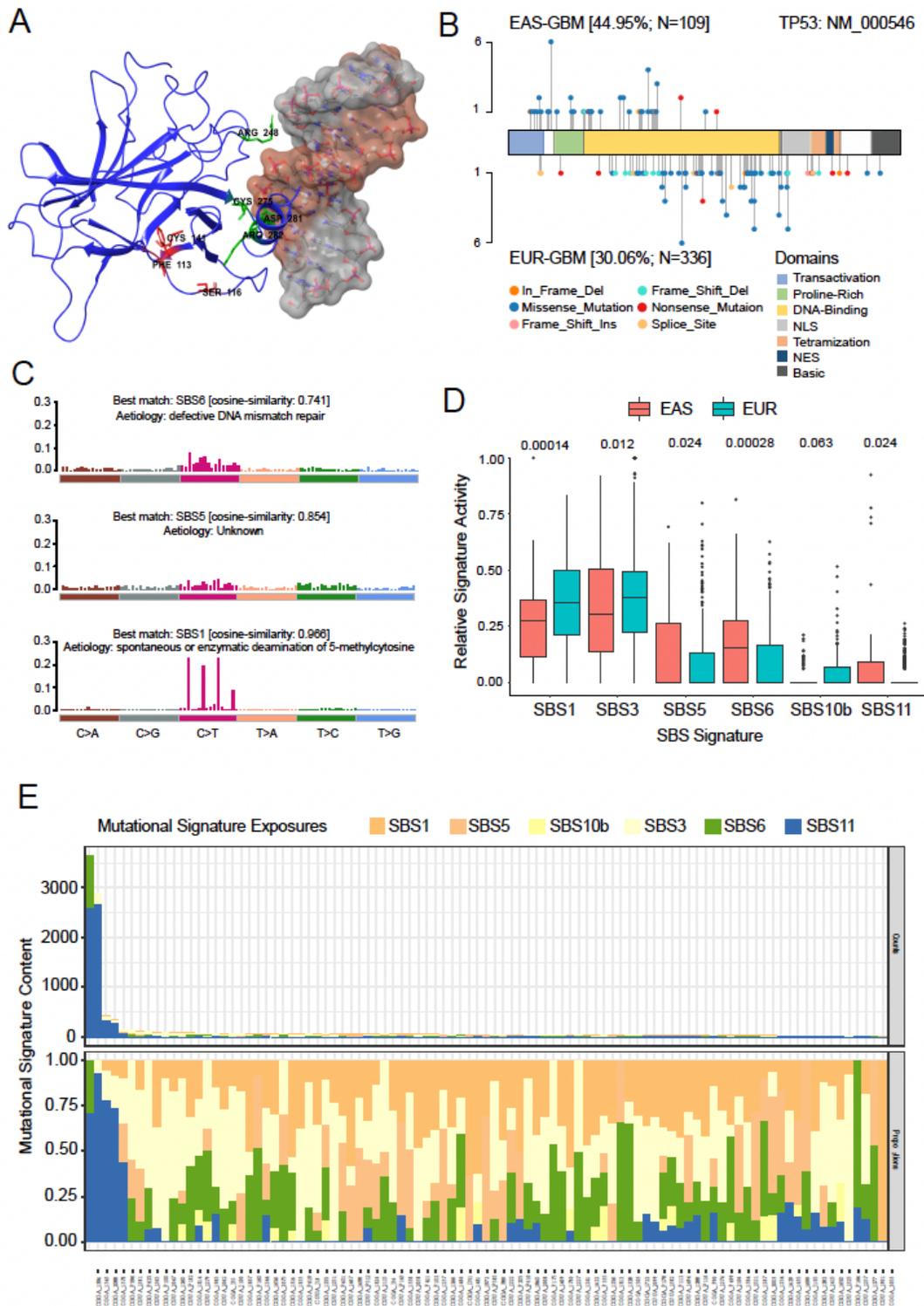


Figure 2

(A) Variant positions in EAS-GBM (red) and EUR-GBM (green) projected on the crystal structure of human P53 protein. (B) Positions and frequency of TP53 somatic variants in EAS-GBM (top) and EUR-GBMs

(bottom) projected on the TP53 domain structure. (C) Mutational signatures (SBS 1, SBS 5, SBS 6) identified in EAS-GBM. (D) Relative mutational signature activity across samples in EAS vs. EUR ancestries. (E) Mutation signatures exposures of all EAS-GBM participants.

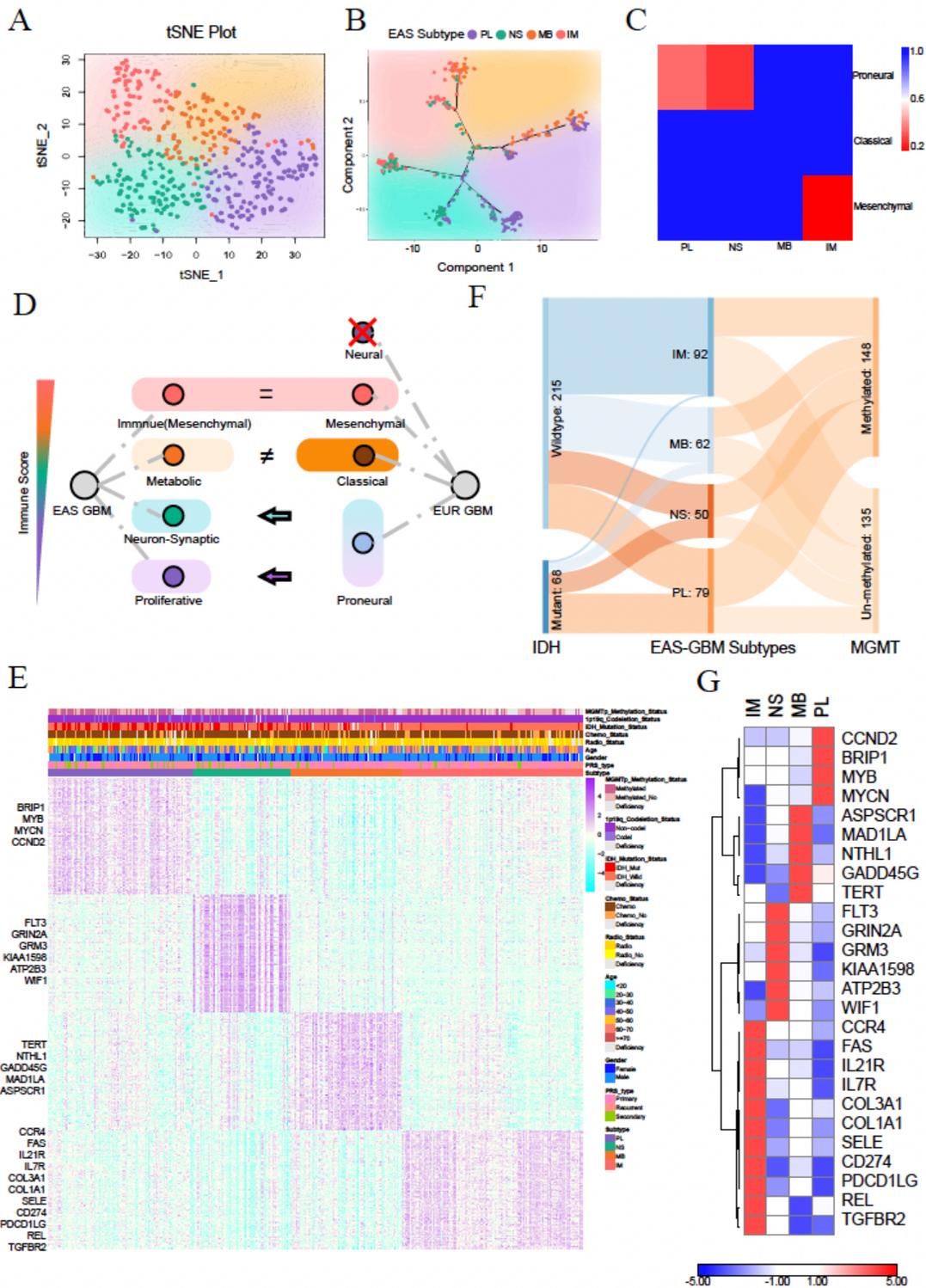


Figure 3

(A) t-SNE plot of the EAS-GBM samples shows clustering by NMF defined groups. (B) Phylogenetic tree plot showed that the genetic distance between EAS-PL GBM and EAS-IM GBM are much larger than

others. (C) SubMap results showed that EAS-PL and NS GBM both correspond to EUR Proneural GBM, EAS-IM GBM are consistent with EUR Mesenchymal GBM. (D) Schematic of GBM subtype alignment across EAS and EUR: Both EAS-PL and NS GBM correspond to EUR Proneural GBM. EAS-MB GBM does not correspond to EUR classical GBM. EAS-IM GBM corresponds to EUR Mesenchymal GBM. (E) Top marker genes accounting for each EAS GBM subtypes were screened by Comparative Marker Selection algorithm. (F) The correlation between EAS-GBM molecular subtypes and IDH mutation status and MGMT promoter methylation. (G) Significance of expression enrichment for selected marker genes of each EAS-GBM subtype.

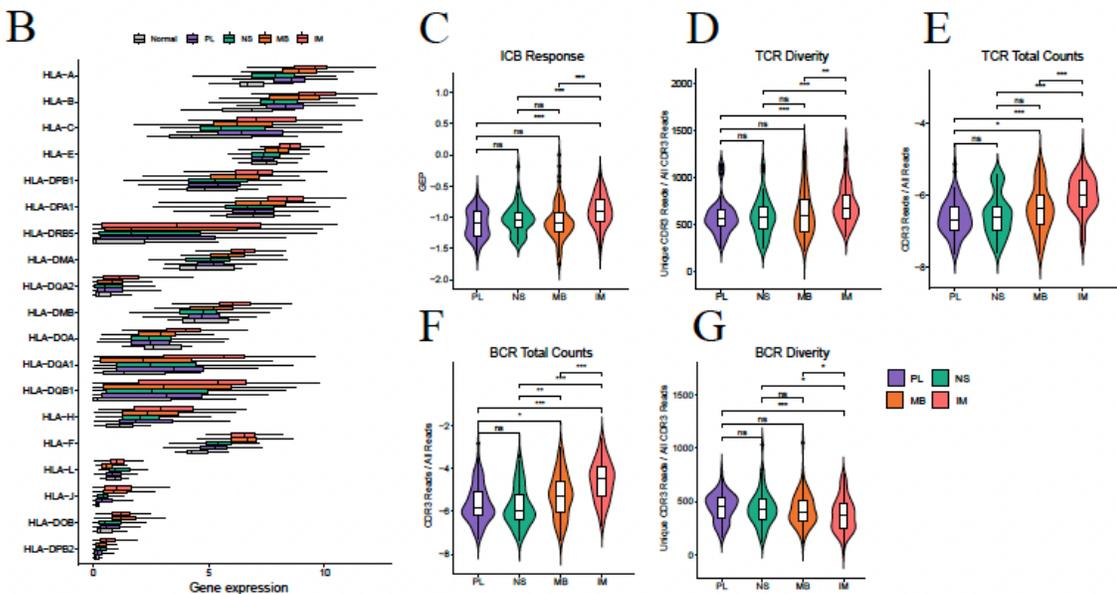
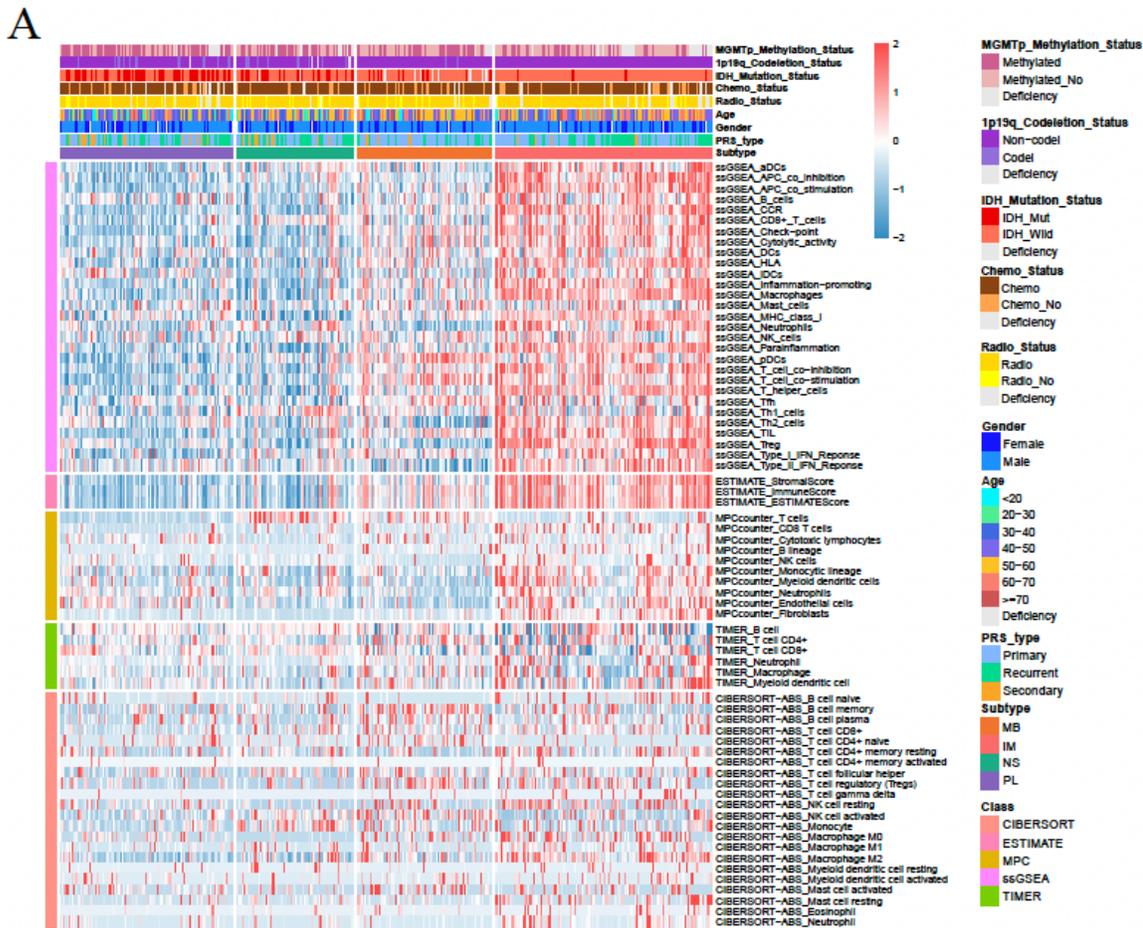


Figure 4

(A) Immune infiltration heatmap of EAS-GBM. The clinical, molecular information and genomic phenotypes of EAS GBMs are shown at the top nine rows with their EAS-GBM subtype label. The following rows show the values of immune-related signatures. Red denotes intense immune infiltration and blue represents low immune infiltration. (B) HLA gene expression among the four EAS-GBM clusters. (C) IFN- γ / T-cell-inflamed gene expression profile (GEP) score of four clusters of EAS GBMs is evaluated

to predict ICB therapy response. (D) Unique TCR CDR3 counts per kilo reads (CPK) approximating TCR clonal diversity among four clusters of EAS GBMs. (E) The total counts of TCR CDR3s approximating the amount of T-cell infiltration among four clusters of EAS GBMs. (F) The total counts of BCR CDR3s approximating the amount of B-cell infiltration among four clusters of EAS GBMs. (*p value < 0.05, **p value < 0.01, ***p value < 0.001) (G) Unique BCR CDR3 counts per kilo reads (CPK) approximating BCR clonal diversity among four clusters of EAS GBMs.

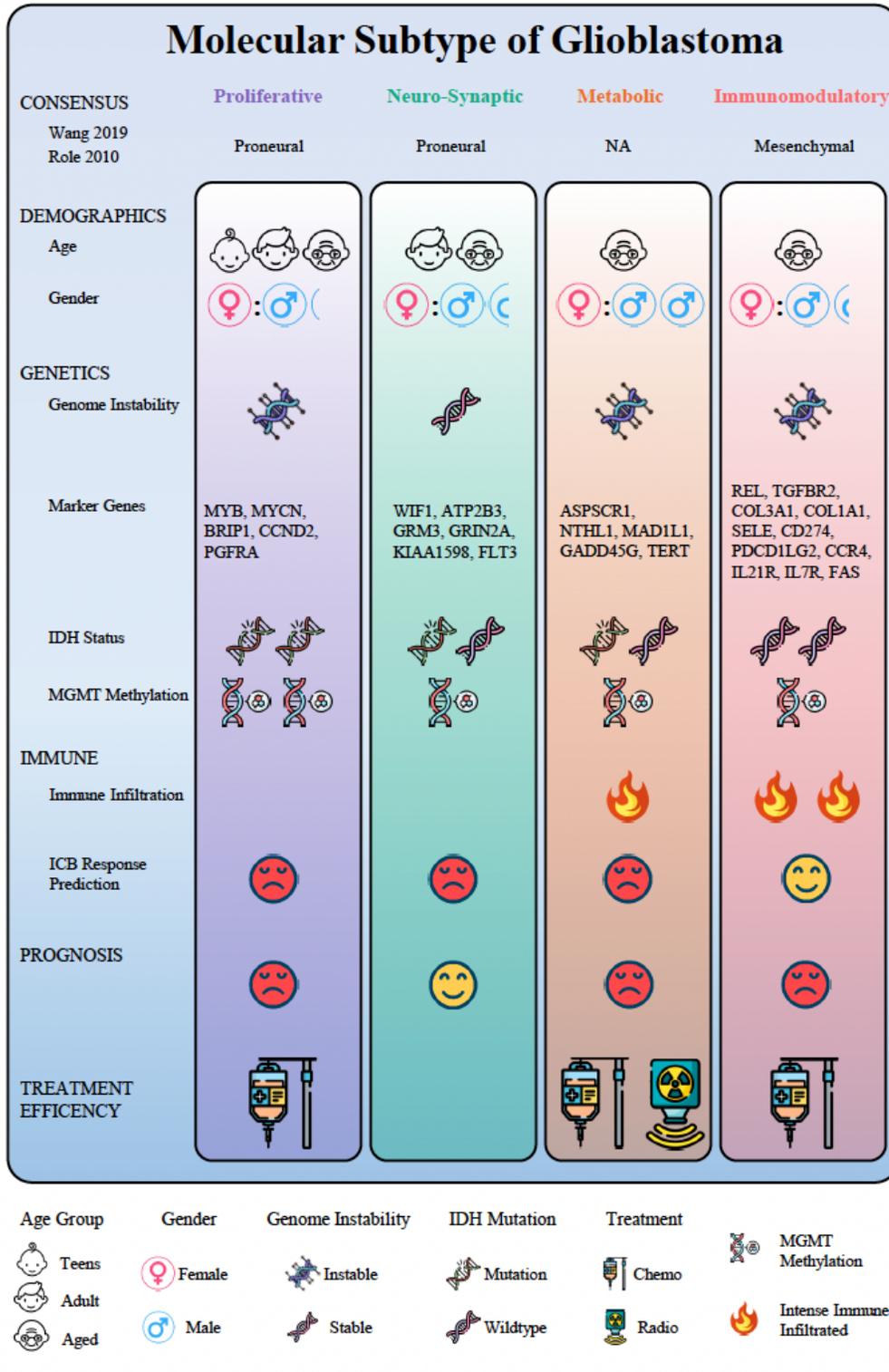


Figure 5

Genomic and clinical characteristics of the four EAS GBM subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ExtendedMaterials.pdf](#)
- [ExtendedFigure1.pdf](#)
- [ExtendedFigure2.pdf](#)
- [ExtendedFigure3.pdf](#)
- [ExtendedFigure4.pdf](#)
- [ExtendedTable1.xlsx](#)
- [ExtendedTable2.xlsx](#)
- [ExtendedTable3.xlsx](#)
- [ExtendedTable4.xlsx](#)