Neuro-Oncology
21(2), 201–213, 2019 | doi:10.1093/neuonc/noy154 | Advance Access date 25 September 2018

Sex difference of mutation clonality in diffuse glioma evolution

Hongyi Zhang,† Jianlong Liao,† Xinxin Zhang,† Erjie Zhao,† Xin Liang, Shangyi Luo, Jian Shi, Fulong Yu, Jinyuan Xu, Weitao Shen, Yixue Li, Yun Xiao, and Xia Li

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China (H.Z., J.L., X.Z., E.Z., X.L., S.L., J.S., F.Y., J.X., W.S., Y.L., Y.X., X.L.)

Corresponding Authors: Xia Li, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, 194 Xuefu Road, Harbin 150081, China 150081 (lixia@hrbmu.edu.cn), and Yun Xiao, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, 194 Xuefu Road, Harbin 150081, China 150081 (xiaoyun@ems.hrbmu.edu.cn).

†These authors contributed equally to this work.

Abstract

Background. Sex differences in glioma incidence and outcome have been previously reported but remain poorly understood. Many sex differences that affect the cancer risk were thought to be associated with cancer evolution.

Methods. In this study, we used an integrated framework to infer the timing and clonal status of mutations in ~600 diffuse gliomas from The Cancer Genome Atlas (TCGA) including glioblastomas (GBMs) and low-grade gliomas (LGGs), and investigated the sex difference of mutation clonality.

Results. We observed higher overall and subclonal mutation burden in female patients with different grades of gliomas, which could be largely explained by the mutations of the X chromosome. Some well-established drivers were identified showing sex-biased clonality, such as *CDH18* and *ATRX*. Focusing on glioma subtypes, we further found a higher subclonal mutation burden in females than males in the majority of glioma subtypes, and observed opposite clonal tendency of several drivers between male and female patients in a specific subtype. Moreover, analysis of clinically actionable genes revealed that mutations in genes of the mitogen-activated protein kinase (MAPK) signaling pathway were more likely to be clonal in female patients with GBM, whereas mutations in genes involved in the receptor tyrosine kinase signaling pathway were more likely to be clonal in male patients with LGG.

Conclusions. The patients with diffuse glioma showed sex-biased mutation clonality (eg, different subclonal mutation number and different clonal tendency of cancer genes), highlighting the need to consider sex as an important variable for improving glioma therapy and clinical care.

Key Points

1. Comprehensive analysis of mutation clonality in ~600 diffuse gliomas
2. Higher overall and subclonal mutation burden in female gliomas
3. Sex difference of clonal tendency for important driver genes

Diffuse glioma, including low-grade glioma (LGG) and glioblastoma (GBM), is a common primary malignant intracranial tumor in adults. It accounts for almost 80% of malignant brain tumors and has high mortality, especially in GBM. According to previous studies, sex differences present in both incidence and outcome of glioma patients, with higher morbidity and mortality in men than women.[1,2] Although the obvious epidemiological disparity exists between male and female glioma patients, neither pathological diagnosis nor clinical treatment considers sex as an important variable.

Sexual dimorphisms of glioma at the clinical phenotypic and molecular levels have been revealed by several researches. For example, in a retrospective study

Importance of the study

Many sex differences shown in brain tumor were thought to be secondary products during evolutionary processes. However, the association between sex and evolution of glioma and the impact of sex on clonal evolution are still poorly investigated. In this study, we used the genomic data of diffuse glioma (GBM and LGG) from TCGA to infer the timing and clonal status of mutations for each sample. We demonstrated, for the first time, sex difference of mutation clonality in glioma genome evolution, highlighting the need to consider sex as an important variable for improving glioma therapy and clinical care.

of malignant glioma patients treated by temozolomide, researchers found that female patients had a significantly higher grade 3/4 myelotoxicity compared with males.[3] Sun et al found that intrinsic differences in *RB* activation between sexes may underlie the predominance of mesenchymal glioblastoma in males.[4]In addition, a recent study reported that the genetic polymorphisms of *ADCY8* could influence the glioma risk through a sex-specific manner in type I neurofibromatosis patients.[5] These works dissected the sex differences from the aspect of individual genes. However, an effort to systematically identify the genomic alterations associated with sex difference in glioma is still lacking.

Many sex differences shown in brain tumor were thought to be secondary products during evolutionary processes.[6] In the last few years, several high-quality works successfully utilized mutation data to characterize the clonal evolution of glioma.[7,8] These studies did not investigate the association between sex and evolution of glioma. The impact of sex on clonal evolution is still poorly understood. Throughout men's and women's lives, there are many sex-specific differences in exogenous factors, such as daily behavior and intrinsic factors such as hormonal milieu, some of which can lead to different microenvironments and selective pressure for tumor cells, causing distinct mutation patterns during genome evolution of cancer. It is thus reasonable to suppose that there should be sex-specific mutation patterns during clonal evolution of glioma. At present, with the availability of high-throughput genomic data in large-scale genomic projects such as The Cancer Genome Atlas (TCGA) Network, we are able to dissect mutation patterns in glioma evolution and further reveal sex differences in mutation clonality, which can facilitate our understanding of sex dimorphism in phenotype and clinical outcome.

In this study, we used the genomic data of diffuse glioma (GBM and LGG) from TCGA to infer the clonal status of mutation for each sample. We characterized the difference of mutation clonality between male and female patients. Besides, based on clonal status of mutations, we identified sex-biased clinically actionable driver genes and pathways, suggesting the clinical importance of considering sex in the treatment of gliomas.

Materials and Methods

Data Source

Our data included exome sequencing data (level 2, $n = 803$), Affymetrix SNP6 data (level 3, $n = 1090$), and clinical data ($n = 1105$) of patients with GBM and LGG from TCGA, in which 791 glioma samples were measured with both somatic mutation and copy number data. The mutation and copy number data of LGG and the copy number data of GBM were acquired from TCGA or Broad Institute Firehose (https://gdac.broadinstitute.org). The mutation data of GBM including variant allele frequencies of mutations were obtained from cBioPortal (http://www.cbioportal.org). The detailed information is shown in Supplementary Table 1. The GBM and LGG driver genes used in this study refer to those recurrently mutated genes that were identified using the MutSig algorithm (Supplementary Tables 2 and 3). We collected these drivers from 2 recent TCGA studies.[9,10] The clinically actionable genes, which show therapeutic relevance and mutations which have therapeutic or prognostic implications, were obtained from the TARGET v2 database (www.broadinstitue.org/cancer/cga/target) for evaluating mutation clonality.

Inferring the Cancer Cell Fraction and Clonality of Somatic Mutation

The cancer cell fraction (CCF) of each mutation was estimated by integrating tumor purity and local copy number as outlined by McGranahan et al[11] and Landau et al.[12] First, we used the mutation data of exome-seq (level 2) and copy number data of SNP6 array (level 3) as input of the ABSOLUTE,[13] a computational algorithm that allowed us to obtain the tumor purity and the absolute DNA copy number for each glioma sample (Fig. 1). Next, we extracted the local copy number for each mutation site from the ABSOLUTE results. The expected variant allele frequency (VAF) of each mutation site depends on the purity (p), the local copy number of this mutation site CPN_{mut}, and CCF. Thus, for a given CCF, the expected VAF can be calculated according to the following equation:

$$VAF_{ex} = \frac{p * CCF * CN_{mut}}{CPN_{norm}(1-p) + p * CPN_{mut}},$$

where CN_{mut} is the mutation copy number (let $CN_{mut} = 1$) and CPN_{norm} denotes the absolute copy number of the normal cell. We generally let $CPN_{norm} = 2$ (let $CPN_{norm} = 1$ when considering mutations in the X chromosome for males). Therefore, the expected VAF also can be represented as:

$$VAF_{ex} = \frac{p * CCF}{2(1-p) + p * CPN_{mut}}.$$

For a point mutation with a alternate reads and N sequencing coverage, the probability of a given CCF is estimated by using Bayesian probability theory and a binomial distribution:

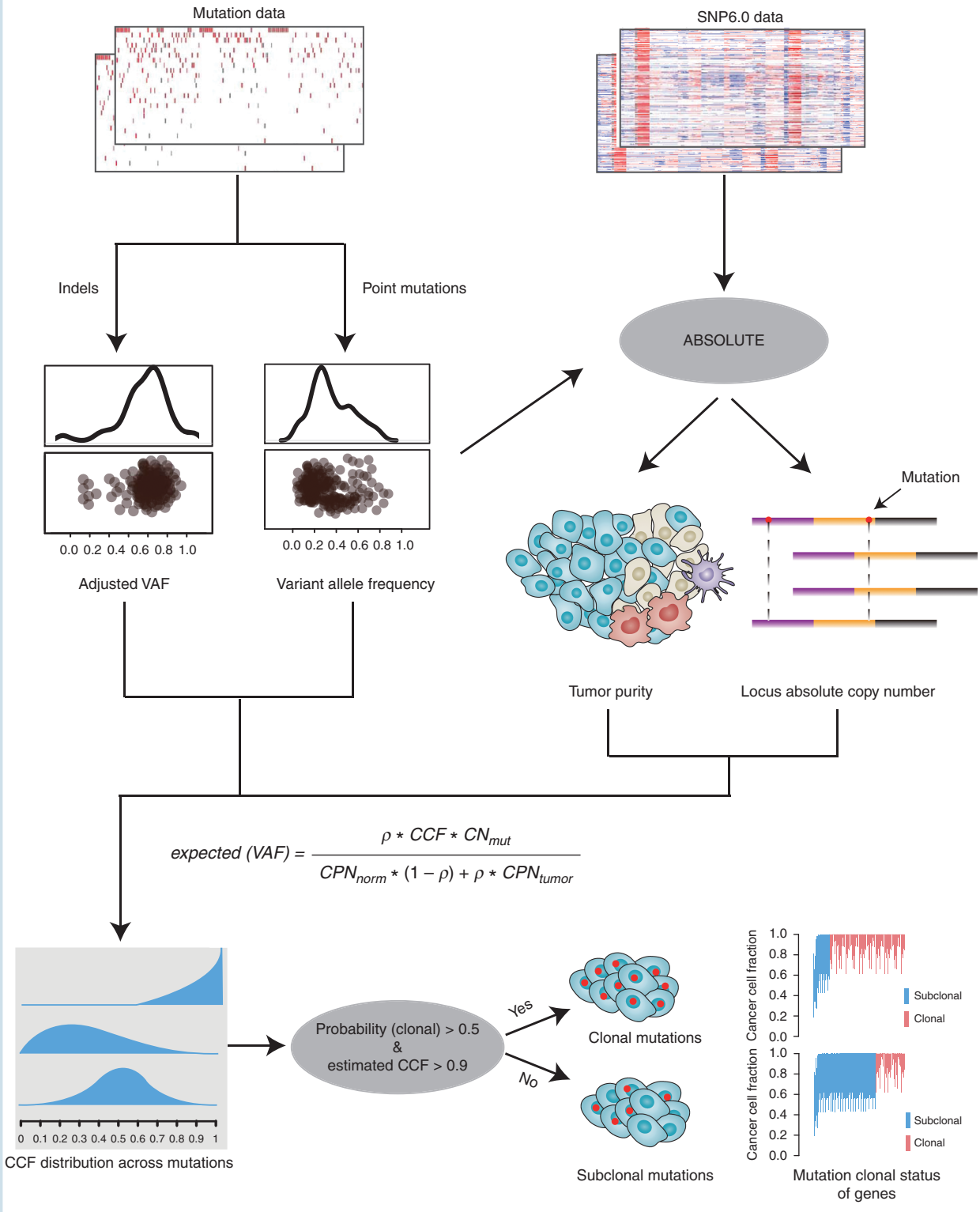

Fig. 1 Workflow of inferring the cancer cell fraction (CCF) and clonality of somatic mutation. Integration of single-nucleotide polymorphism (SNP) arrays and exome sequencing data to infer CCFs of mutations (see Methods section for details).

$$P(CCF \mid (a \mid N)) = \frac{P((a|N)|CCF) * P(CCF)}{P(a \mid N)}$$

and

$$P(CCF \mid (a \mid N)) \propto \text{Binom}(a \mid N,\ VAF_{ex}(CCF)).$$

Then a distribution of CCF was obtained by calculation of the $P(CCF)$ over a uniform grid of 100 CCF values from 0.01 to 1 and normalization by dividing their sum.

The observed VAF of indel mutations must be adjusted when estimating their CCFs, because of the biases affecting the alignment of short reads that always favor reference over alternate alleles, thus leading to a lower VAF than expected.[12] To adjust such biases, we grouped indels based on their lengths and compared the observed VAFs of each group with those of clonal single nucleotide variations in a diploid region of samples with tumor purity >0.75. We found that the VAFs for indel mutations peaked at 0.40, 0.38, 0.35, 0.33, 0.31 for 1-based indels, 2-based indels, 3-based indels, 4-based indels, and ≥5-based indels, respectively (Supplementary Fig. 11). Therefore, the bias factor *BF* could be estimated in each indel group by dividing by the peak of single nucleotide variations. The adjusted VAFs of indel mutations were calculated as:

$$adj(VAF) = \frac{VAF}{BF}.$$

The estimation of CCF values for indel mutations is based on the $adj(VAF)$, and other calculation procedures are analogous to point mutations. The estimated CCF for each mutation is:

$$\text{argmax}_i\left(P(CCF = i)\right).$$

In order to determine the clonal status of each mutation, we defined the $Pr(clonal)$ representing the probability that a mutation is clonal:

$$Pr(clonal) = P(CCF \geq 0.9) = \sum_{i=0.9}^{1} P(CCF = i).$$

Finally, mutations were classified as clonal if the estimated CCF was >0.9 and the $Pr(clonal)$ was >0.5, and as subclonal otherwise.

Notably, the indel data downloaded from cBioPortal have filtered out indels with low VAFs (<10%). To examine whether these removed indels (~2% of all mutations) can influence our results, we additionally obtained the mutation files without removing those low-VAF indels for 465 glioma samples used in this study from Firehose. We observed the same results that female patients have higher overall and subclonal mutation burden than males, in both GBM and LGG ($P < 0.05$, Wilcoxon rank-sum test) when including these removed indels. These suggest that the indels removed by cBioPortal were unlikely to have a major effect on our results.

Statistical Analysis

The Wilcoxon rank-sum test was used to compare mutation burden, clonal mutation burden, and subclonal mutation burden between males and females. Moreover, we validated the statistical significance by a permutation test (ie, randomly shuffling sex labels of patients) (Supplementary Table 8). We applied Fisher's exact test to analyze sex differences of driver gene mutation frequency and clonality. The sex difference of clonal status for clinically actionable genes was also analyzed by Fisher's exact test. To establish the clonal/subclonal preference of driver genes in males or females, we compared the clonal fraction between nonsilent mutations and background silent mutations in each driver using a binomial test. Survival analyses were conducted with the use of the Kaplan–Meier and Cox proportional hazards regression methods, and survival distributions were compared across groups with the use of the log-rank test. More detailed information on the methods used in this study can be found in the Supplementary Methods.

Results

Sex-Biased Overall Mutation Burden in Gliomas

To determine if there was a sexual dimorphism in glioma mutation data, we obtained the mutation spectra of all diffuse glioma samples in TCGA including GBM (295 samples) and LGG (508 samples). We noted that TCGA contained sufficient glioma samples with different sex for mutation data, which enabled us to analyze sex differences in cancer genomes (male = 187 and female = 108 in GBM, male = 282 and female = 226 in LGG, Supplementary Table 1). A total of 30 228 somatic mutations were obtained, including 19 428 missense, 7259 silent, 1328 nonsense, 15 nonstop, 1101 frame-shift indel, 315 in-frame indel, 16 intron, 33 noncoding region (5′ untranslated region [UTR], 3′ UTR, 5′ flank, 3′ flank or RNA), 701 splice site or region, and 32 translation start site mutations.

We found a statistically significantly higher burden of mutations in females than in males (GBM: male median = 49 vs female median = 56.5, $P = 0.00011$; LGG: male median = 24 vs female median = 26, $P = 0.044$; Wilcoxon rank-sum test) (Supplementary Fig. 1A). Such higher mutation burden in female patients could also be observed in GBM when removing the silent and noncoding mutations (Supplementary Fig. 1B). We further tested how individual chromosomes contribute to sex disparity of overall mutation burden. Through comparing the mutation burden for each chromosome between males and females, we found that 4 chromosomes in GBM showed higher mutation numbers in females, including chromosomes 1, 14, 21, and X (Benjamin–Hochberg false discovery rate [FDR] < 0.05; Fig. 2A). Besides, chromosomes 3 and 8 also showed marginal sex differences of mutation burden ($P < 0.05$, FDR < 0.1). In LGG, only the X chromosome showed a significant sex difference after adjusting for multiple testing ($P < 0.01$, FDR < 0.05; Fig. 2B). These results indicated that X chromosomes contributed a lot to the sex difference of the overall mutation load (especially in LGG), consistent with a previous study in which sex-biased genes were revealed to be enriched in sex chromosomes.[14] To determine whether the sex difference

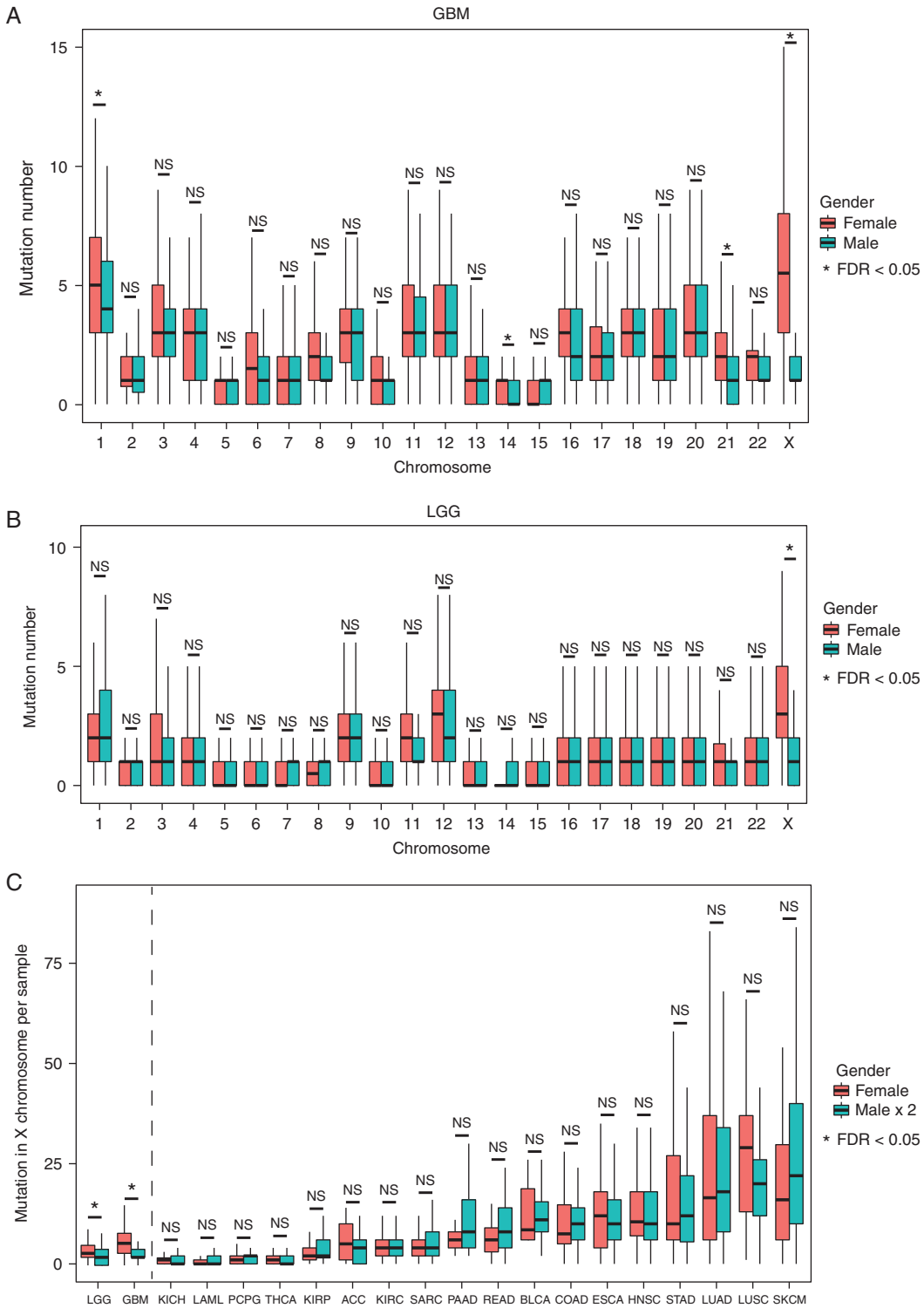

Fig. 2 Mutational burden comparison between males and females across different chromosomes. (A) Mutational burden comparison between sexes in each chromosome of GBM. (B) Mutational burden comparison between sexes in each chromosome of LGG. (C) Mutational burden comparison between sexes in X chromosomes across 20 cancer types.

in mutation numbers of X chromosomes was glioma specific or just due to the X chromosome haploidy in males, we compared mutation load of X chromosomes between males and females in 18 other TCGA cancer types (Fig. 2C). Before comparison, we simply doubled the mutation number for each male patient to adjust the effect of X chromosome haploidy in males. As a result, among 20 cancer types, only GBM and LGG were found to exhibit statistically significant sex differences in mutation burden ($P < 0.001$, FDR < 0.05). In summary, the sex disparity in mutation burden of X chromosomes is specific for gliomas, which further contributed to the sex difference in overall mutation burden.

To further explore the sex differences at the single gene level, we identified all nonsilent mutations that occurred in driver genes derived from 2 recent TCGA works[9,10] and tried to identify sex-biased driver genes. Unfortunately, we found no driver gene showing difference in mutation frequency between male and female tumors, consistent with the study by Yuan et al, which also failed to identify sex-biased mutated genes in GBM and LGG.[14]

Inference of Mutation Clonality in a Single Glioma Sample

From previous works, we knew that differences in brain tumor risk and outcome are evident between males and females and such sex differences may be by-products during evolutionary processes.[6] Actually, we found significant sex differences of mutation burden that may reflect a sex-biased selective pressure leading to mutagenesis in glioma clonal evolution. In order to investigate sex differences of clonal evolution in gliomas, we thus inferred the clonal status of mutations and performed sex-specific analyses across GBM and LGG. We conducted an integrative analysis to infer the CCF of each mutation (the fraction of tumor cells carrying this mutation within a sequencing sample) using single nucleotide polymorphism (SNP) arrays and exome sequencing data (Fig. 1). Mutations were classified as clonal if the estimated CCF was >0.9 and the *Pr*(*clonal*) was >0.5, and as subclonal otherwise. Notably, we calculated the detection power for all mutations with various VAFs and removed those mutations with power <80% (see Supplementary Methods).

Among the 803 TCGA glioma samples with exome-seq data, the tumor purity and local absolute copy numbers of 591 samples were calculated using ABSOLUTE (GBM: 262 and LGG: 329, with the least purity >20%). The other samples with low purity and heavily contaminated tumors were not used for clonal analysis. Totally, 10 498 (37.7%) clonal mutations (GBM: 6074, LGG: 4424) and 17 349 subclonal mutations (GBM: 9713, LGG: 7636) were identified, significantly overlapping with the results of Kim et al ($P < 2.2 \times 10^{-22}$, chi-square test),[15] who also performed analysis of mutation clonality across the TCGA GBM dataset. Clonal mutations presenting in all tumor cells represent the early events in tumor evolution, because they are likely to occur before or during the most recent complete selective sweep. In contrast, subclonal mutations tend to occur after the emergence of the most recent common ancestor, hence representing relatively late events.[16]

Clonal Difference in Glioma Evolution Between Male and Female Patients

Based on the inferred clonal status of mutations, we compared the clonal mutation burden and subclonal mutation burden between males and females in GBM and LGG. To our surprise, we found significant differences in subclonal mutation burden, but not in clonal mutation burden (Supplementary Fig. 2A). Female patients carried higher subclonal mutation numbers than male patients in both GBM and LGG (GBM: male median = 33.5 vs female median = 38, $P = 0.00168$; LGG: male median = 19 vs female median = 21, $P = 0.017$). The odds ratios for subclonal mutations in female patients were calculated as 1.09 (95% CI: 1.02–1.16) and 1.18 (95% CI: 1.10–1.28) for GBM and LGG, respectively, indicating that sex-biased overall mutation burden was dominated by subclonal mutations. When focusing on nonsilent mutations, consistent observations were obtained (Supplementary Fig. 2B). Next, we wondered if the mutations of the X chromosome have a main effect on such sex difference. Similar to the observation in analysis of overall mutation burden, we found that mutations in the X chromosome made a major contribution to the sex difference (Supplementary Fig. 3).

To determine whether similar results present in patients with different World Health Organization (WHO) grades (II, III, and IV) and histologic subtypes (oligodendroglioma, astrocytoma, and glioblastoma), we performed mutation burden analysis in each grade and each histologic subtype. We observed higher subclonal mutation burden of females in grades III and IV, astrocytoma, and glioblastoma samples (Supplementary Fig. 2C, D), suggesting that the sex-biased subclonal mutation burden was associated with more malignant gliomas. Since the new version of WHO guidelines include isocitrate dehydrogenase (*IDH*) mutation (status of glioma cytosine-phosphate-guanine island methylator phenotype [G-CIMP]) and 1p/19q codeletion as the key drivers of diffuse glioma,[17] we also tested whether sex-biased subclonal mutation load existed in glioma subgroups stratified based on *IDH* (*IDH1*/*IDH2*) mutation and 1p/19q codeletion. As a result, higher subclonal mutation burden could be found in female patients of *IDH* wild-type (non–G-CIMP) and *IDH*-mutation only (no 1p/19q codeletion) subgroups (Fig. 3A).

Other factors, such as age and race, may bias our findings. In many cancers, mutation count increases with age.[18] Similarly, we observed that older patients carried more subclonal mutations regardless of sex (Fig. 3B). The association between the subclonal mutation burden and age was not affected by the X chromosome (Supplementary Fig. 4A). To correct the effects of age and other potentially confounding factors (Supplementary Fig. 4A, B), we used the propensity score for balancing the confounding factors in the 2 sex groups[19] (see Supplementary Methods). We estimated the propensity scores using age at diagnosis, tumor purity, race, *IDH1*/*IDH2* mutation (G-CIMP status), 1p/19q codeletion, WHO grade, and histologic subtype as covariates and then obtained a balanced patient cohort based on the estimated propensity scores. This step removed 118 male samples (GBM: 69 [42.9%], LGG:

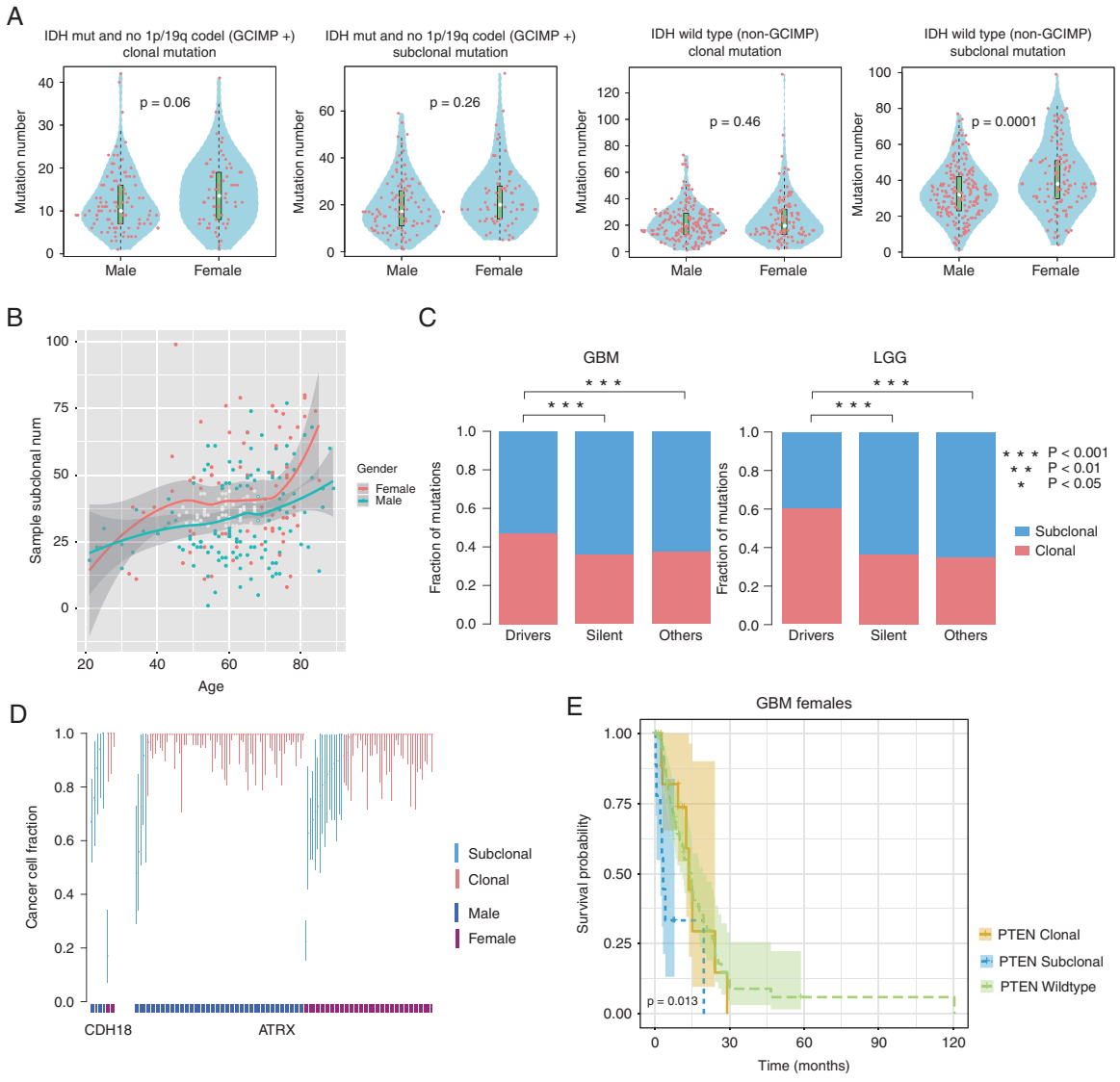

Fig. 3 Clonal differences between males and females with GBM or LGG. (A) Comparison of clonal and subclonal mutation burden between 2 sex groups across glioma molecular subtypes. Significance from Wilcoxon rank-sum test is indicated. (B) The correlation of subclonal mutation number and age in patients with GBM. The red dot represents female patient and the blue dot represents male patient. The fitted curves reflect the changing tendency of subclonal mutation number with age. (C) The clonal fraction of aggregated driver mutations versus silent mutations and other nondriver mutations. Clonal mutations are marked as red and blue represents subclonal mutations. Significance level of Fisher's exact test is also indicated. (D) The CCFs of mutations in *CDH18* and *ATRX* which show a sex-biased clonal tendency in GBM or LGG are shown. Each symbol represents a somatic mutation in a single tumor sample marked as male or female using a bottom bar colored by dark blue and dark red, respectively. Red point represents the estimated CCF of clonal mutation and blue point represents the estimated CCF of subclonal mutation. Error bars represent the 95% confidence interval. (E) Kaplan–Meier estimates of overall survival in 3 female groups with GBM. Orange curve represents samples carrying *PTEN* clonal mutations. Blue curve represents samples carrying *PTEN* subclonal mutations. Green curve represents samples without *PTEN* mutations.

49 [28.8%]; Supplementary Tables 4–7, Supplementary Fig. 5A–D). We next compared the clonal and subclonal mutation burden between these 2 balanced sex groups. As expected, we found a significantly higher subclonal mutation burden in female patients after adjusting for the confounders (GBM: $P = 0.0071$, LGG: $P = 0.023$, Wilcoxon rank-sum test; Supplementary Fig. 5F and H), indicating that the association between sex and subclonal mutations was independent of other factors in brain gliomas. Notably, for clonal mutation burden, we could not observe any sex differences in both GBM and LGG (Supplementary Fig. 5E and G).

The identification of clonal status enables us to analyze the sex difference for mutation patterns of glioma driver

genes. We observed, in both male and female patients, a clear tendency for mutations in driver genes to be clonal compared with background silent mutations and mutations in nondriver genes across GBM and LGG (Fig. 3C), supporting that driver genes preferentially acquire mutations in the early stage of cancer evolution and contribute to gliomagenesis without sex bias (Supplementary Fig. 6). Further, we focused on those recurrent drivers mutated in at least 10 samples of GBM or LGG and sought to reveal whether mutations of these driver genes tend to be clonal in a specific sex group by comparison to background silent mutations. In GBM, we identified 6 driver genes showing the tendency to be clonal in males and 3 genes in females ($P < 0.05$, FDR < 0.1; Supplementary Table 9). Among these genes, *ATRX*, *PDGFRA*, and *STAG2* showed clonal tendency only in males (Supplementary Fig. 7A). Similarly, in both sexes of LGG, we found 4 genes showing higher clonal fraction than silent mutations ($P < 0.05$, FDR < 0.1; Supplementary Table 10). In these genes, *EGFR* showed enrichment of clonal mutations only in males, whereas *ZBTB20* showed enrichment only in females (Supplementary Fig. 7B). Furthermore, we directly compared clonal fractions between different sexes using Fisher's exact test. We observed that the clonal fraction of *CDH18* mutations showed an opposite tendency in GBM (clonal fraction in male vs female = 0% vs 75%, $P < 0.05$, FDR < 0.1; Fig. 3D). Also, we found 2 genes showing differential clonal fraction between the 2 sex groups of LGG (*ZBTB20*, clonal fraction in male vs female = 16.7% vs 100%, $P < 0.05$, FDR < 0.1; *ATRX*, 93.0% vs 70.4%, $P < 0.05$, FDR < 0.1; Fig. 3D). These results suggested that some driver genes, exclusively showing clonal tendency in a specific sex group, may play distinct oncogenic roles in glioma evolution between males and females.

We next explored whether the mutation clonal status of driver genes could impact patient clinical outcome in a sex-specific manner. We performed survival analyses for each driver gene and found that the patients harboring subclonal mutation of phosphatase and tensin homolog (*PTEN*) had a poorer survival compared with female patients with clonal mutation of *PTEN* in GBM (Fig. 3E and Supplementary Fig. 8). When using multiple-testing adjustment, only a borderline significance was obtained (Benjamin–Hochberg FDR = 0.064), suggesting that a validation cohort is required. A multivariate Cox model incorporating age, race, ethnicity, tumor purity, *IDH* (G-CIMP) status, and transcriptome subtypes confirmed an independent association between clonal status of *PTEN* mutation and overall survival ($P = 0.011$, hazard ratio = 4.81, 95% CI = 1.45 to 16.01). Such prognostic power of *PTEN* did not reach statistical significance when combining male and female patients. Furthermore, the association between subclonal mutation in *PTEN* and poorer survival of GBM female patients was validated in an independent cohort ($P = 0.046$; Supplementary Fig. 9, Supplementary Methods). Because of the low sample size of female patients with PTEN mutation, the prognostic value of PTEN mutation clonality should be further validated in large patient cohorts. Together, these findings indicated that the subclonal mutation in *PTEN* in females may contribute to more malignant cancer progression.

Sex-Biased Subclonal Mutations in Glioma Subtypes

Considering the high heterogeneity in brain glioma, previous studies have established 4 transcriptome subtypes (proneural, neural, classical, and mesenchymal) and 3 molecular subtypes (*IDH* mutation and 1p/19q codeletion, *IDH* mutation and no 1p/19q codeletion, and *IDH* wild type) in GBM and LGG, respectively.[10,20] Some sex differences of brain cancer were found to be subtype specific.[4] Thus, it is necessary to investigate sex differences of mutation clonal status in glioma subtypes.

We first explored the sex differences of clonal and subclonal mutation burden in each subtype. Interestingly, the majority of subtypes showed significantly more subclonal mutations in female patients with GBM and LGG, whereas no subtypes showed sex-biased clonal mutation burden (Fig. 4A, B), with 2 exceptions of the mesenchymal subtype in GBM and the 1p/19q codeletion subtype in LGG, which indicated that the observed sex difference of subclonal mutation burden in the whole patient cohort is not contributed by a specific subtype. Moreover, the sex difference of subclonal mutation burden in the X chromosome was found in all subtypes (Supplementary Fig. 10A, B). We could observe significant sex difference of subclonal mutation burden in the neural and classical subtypes of GBM and the *IDH* mutation and no 1p/19q codeletion subtype of LGG when excluding the effect of the X chromosome (Supplementary Fig. 10C, D).

Next, we restricted our analysis to those frequently mutated drivers ($n > 10$) mentioned above. For 542 mutations in 24 driver genes of GBM, we found a clear tendency to be clonal in proneural and neural patients, regardless of sex (proneural: male $P = 0.00034$, female $P = 0.026$; neural: male $P = 0.02$, female $P = 0.043$; Supplementary Fig. 10E). However, in classical and mesenchymal subtypes, clonal tendency in females was observed (classical: $P = 0.034$ and mesenchymal: $P = 0.0017$), while more subclonal mutations were found in male patients (Fig. 4C and Supplementary Fig. 10E). Especially in mesenchymal GBM, the proportion of subclonal mutations was approximately 60% (55/96) in male patients, obviously higher than that in female patients (40%, 23/57, $P = 0.047$, Fisher's exact test). Notably, we identified several important actionable drivers whose mutation clonal status showed sex differences in specific subtypes. For example, *TP53*, *PTEN*, and *NF1* represent viable therapeutic targets in preclinical models. Most nonsilent mutations of these genes were found to be clonal in mesenchymal female patients, but subclonal in mesenchymal male patients. A similar phenomenon could be observed in *EGFR* and *IDH1* in the proneural subtype (Fig. 4C).

In no LGG subtype did we identify any sex differences of overall clonal status for 844 mutations in 14 driver genes (Supplementary Fig. 10F). But in specific subtypes, we observed that the clonal fractions of some important drivers showed the opposite tendency between male and female patients (Fig. 4D). For instance, in LGG patients with *IDH* mutation and 1p/19q codeletion, all *TP53* mutations were found to be clonal in male patients (clonal vs subclonal: 2 vs 0) but subclonal in female patients (clonal vs subclonal: 0 vs 2). While in the other 2 subtypes of LGG

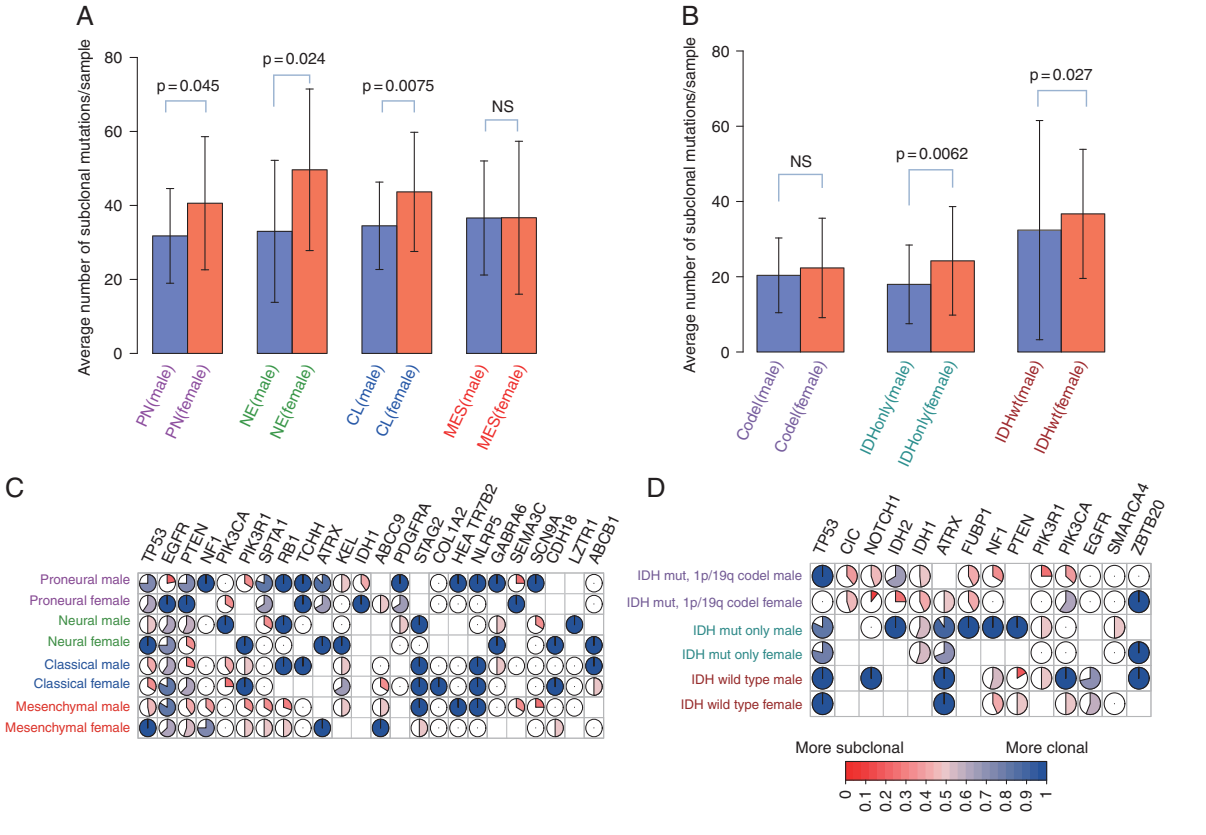

Fig. 4 Comparison of subclonal mutation burden and clonal tendency of driver genes between males and females across glioma subtypes. (A) The number of subclonal mutations in men and women of 4 transcriptome subtypes of GBM. (B) The number of subclonal mutations in men and women of 3 molecular subtypes of LGG. (C) Distribution of clonal fraction of driver gene mutations in males and females of different GBM transcriptome subtypes. Each colored circle sector in the grid chart represents clonal mutation fraction of a driver gene in specific samples. (D) Distribution of clonal fraction of driver gene mutations in males and females of different LGG molecular subtypes.

(*IDH* mutation and no 1p/19q codeletion, *IDH* wild type), *TP53* mutations were preferentially clonal in both sex groups (Fig. 4D). Given that clonal status can reflect the relative timing of mutations occurring during gliomagenesis, these results suggest that, in specific subtypes, key cancer genes may play their oncogenic roles in different evolutionary stages (ie, early and late) between male and female patients.

The Sex Differences in Mutation Clonality of Therapeutic Relevant Genes

Many successful targeted therapies target early clonal events that present in every tumor cell.[21] To study the clinical implications of sex-biased mutation clonality, we next analyzed all nonsilent mutations in actionable genes and pathways for which the therapies have been developed or are in development. For instance, the mutations leading to dysfunction of the mitogen-activated protein kinase (MAPK) pathway have raised much concern about inhibitors of this signaling cascade.[22]

We found 69 actionable genes mutated in GBM and 64 genes mutated in LGG. However, almost all actionable genes (GBM: 53/69 and LGG: 54/64) were found to have a subclonal mutation in at least one sample (Fig. 5A). And the clonal status for mutations in these genes showed statistically significant difference between males and females (Fig. 5B). In GBM, there were more subclonal mutations in the male patients than female, with a marginal significance (P = 0.053, Fisher's exact test). But for LGG, we found more subclonal mutations in female patients than male (P = 0.041, Fisher's exact test). The effect of the X chromosome on the sex difference was not analyzed, because only a few mutations of drug-actionable genes fall on the X chromosome. In addition, some well-studied drug targets showed the opposite tendency of clonal status between men and women, such as *IDH1* in GBM and *MET* in LGG. We next focused on the somatic mutations of actionable genes affecting major cancer pathways (Fig. 6). Interestingly, in GBM, we found that nearly all member genes of the MAPK pathway had higher fractions of clonal mutations in females than males, implying that female patients with GBM may have favorable responses to MAPK inhibitors, such as inhibitors of MEK (mitogen-activated protein kinase kinase). In LGG, we observed higher fractions of clonal mutations in males for the member genes of an important cancer pathway—that of receptor tyrosine

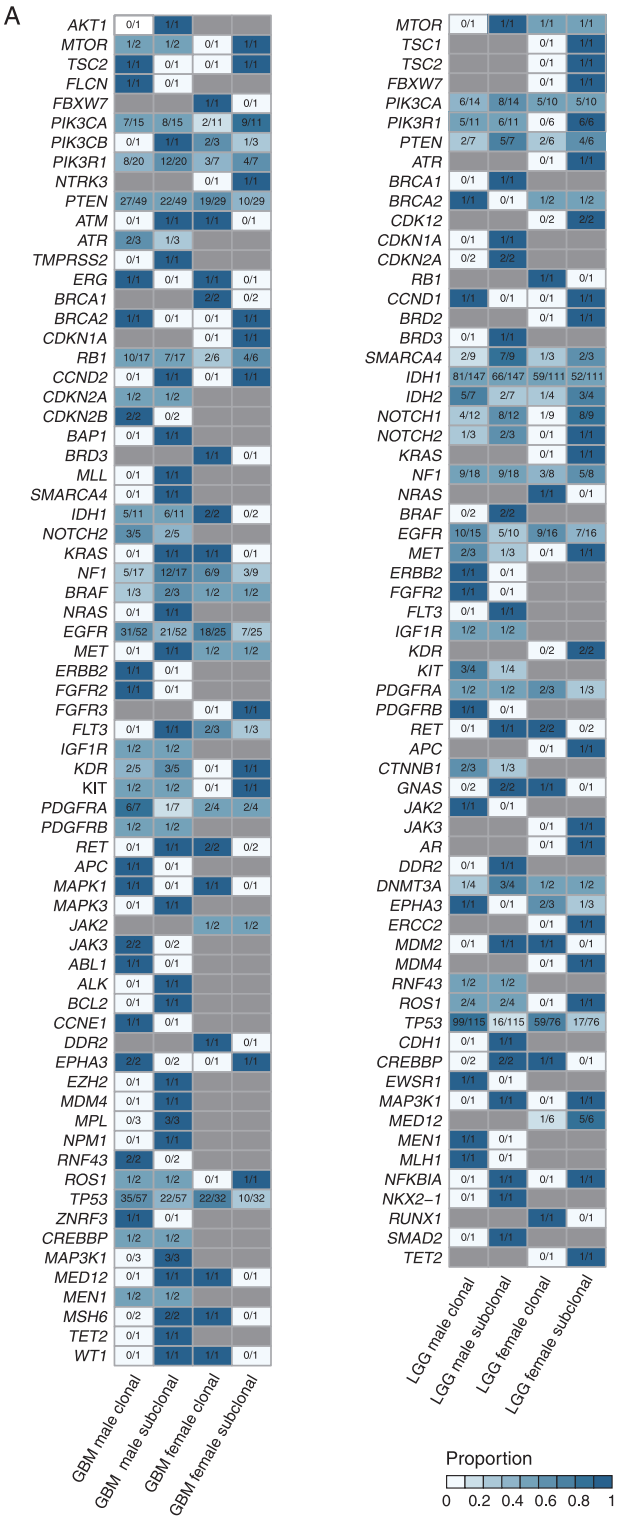

Fig. 5 Sex dimorphism in clonal status of treatment-related genes. (A) Heatmap showing the proportion of nonsilent mutations that are clonal or subclonal for clinically actionable genes in GBM (left) and LGG (right). For each gene, the number of subclonal mutations and clonal mutations is indicated for each sex group. Gray indicates the absence of mutation. (B) The proportion of clonal mutations and subclonal mutations of actionable drivers in males versus females. The *P*-values are calculated from Fisher's exact test to assess statistical significance.

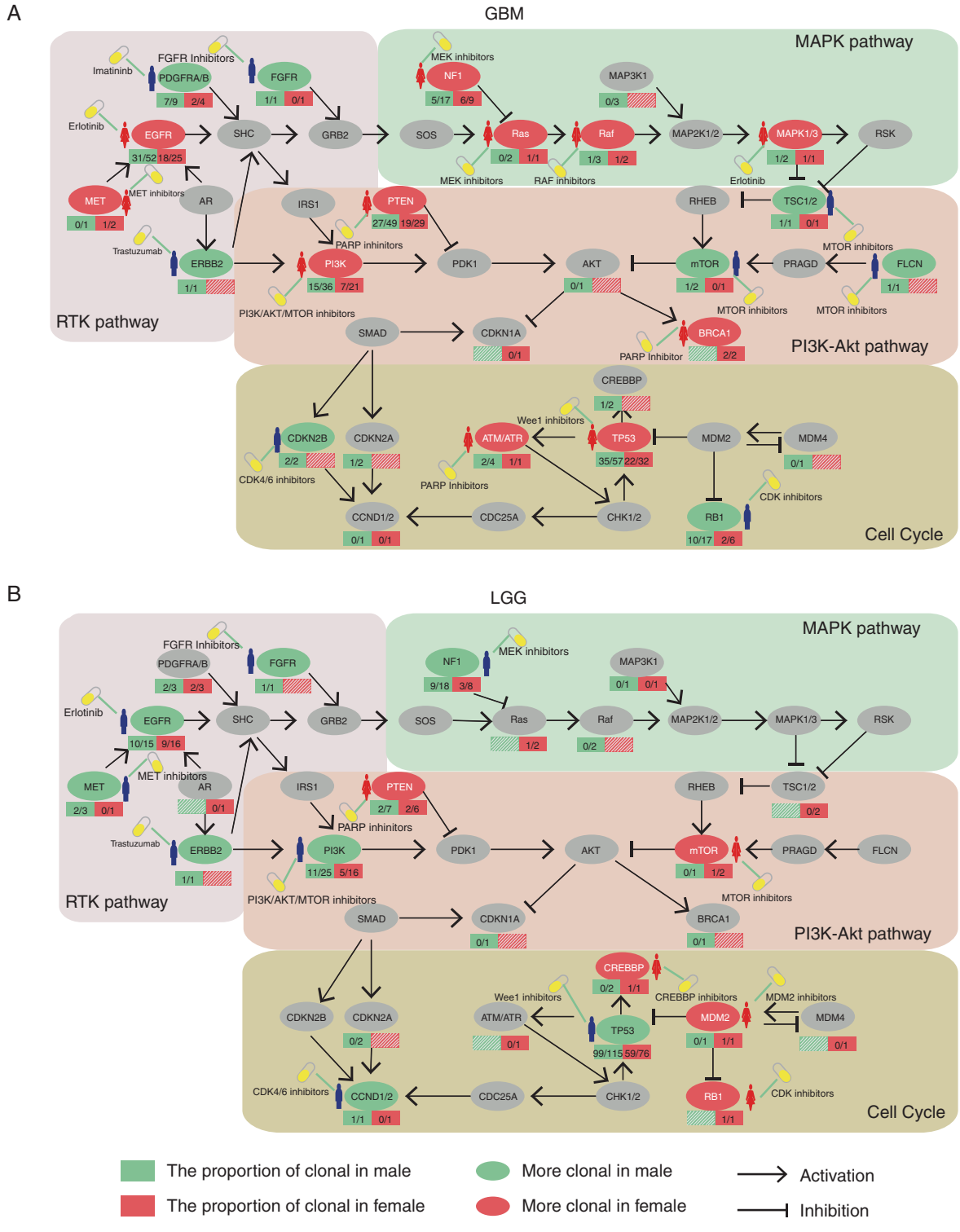

Fig. 6 Sex dimorphism in mutation clonality of treatment-related pathways. Related cancer pathways are depicted for GBM (A) and LGG (B). The number in green and red rectangles denotes the proportion of clonal mutations in males and females, respectively. Each actionable gene showing sexual bias corresponds to specific targeted drugs was marked with color. The red and green filled colors in ellipse indicate that the gene is more clonal in female or male.

kinase (RTK). Together, our findings highlighted the need to take sex-biased treatment and clinical care for a part of glioma patients carrying mutations in specific genes or pathways.

Discussion

The sex effect in tumor progression is widely concerned and recognized as clinically significant.[23] Our study demonstrated, for the first time, sex difference in the genome evolution of glioma and identified some key actionable drivers and pathways showing a sex-biased mutation clonality. The clonal status of genetic mutation is highly correlated with the possibility of a successful targeted therapy, because drugs against a subclonal driver that is present in only a subset of tumor cells may allow the expansion of wild-type subclones, leading to the acquisition of drug resistance.[21] Therefore, our analyses provide important clinical implications for treatment of glioma patients of different sexes.

Our study reported a statistically significant higher subclonal mutation burden in females than males in glioma. A plausible explanation is that the physiological condition is more unstable in females compared with males, particularly after adolescence, which in turn facilitates the accumulation of late mutations. In general, women have to experience some exclusive events throughout their lifetime—pregnancy, childbirth, and menopause. These experiences sometimes can influence the tumor and its microenvironment within female patients (if they have). In fact, cancer risk increases slightly after pregnancy and higher incidence rates are found in women who delay first pregnancy to later in life, suggesting hormone-associated effects on cancer etiology.[6,24] In contrast, young women are unlikely to suffer these sex-specific experiences, which may thus explain why the sex-biased subclonal burden was more obvious in older and more malignant patients.

The evidence from some previous studies can help us to understand the findings in this study. It has been found that sexual dimorphism is involved in tumor cell biology and brain microenvironment,[25] which may confer different selection pressure on somatic cells of male and female brains, thus leading to a sex-biased mutagenesis in glioma evolution. In our results, we found that mutations in the X chromosome contributed a lot to the sex differences of overall and subclonal mutation burden. In particular, significantly higher mutation load was observed in X chromosomes of female glioma patients. The mosaicism of X chromosome inactivation in females may offer a reasonable explanation. This mosaicism was recently found to be associated with an increased mutation rate in cancer (not in normal tissues),[6,26] although such mosaicism is generally protective against disease.

Currently, the use of therapeutic regimens for glioma patients does not have any explicit distinction between men and women. We found that the mutations affecting genes of the MAPK pathway tend to be clonal in females with GBM, while the mutations in genes of the RTK signaling axis prefer to be clonal in males with LGG. For these cases, treating male and female patients in a similar way may lead to completely different responses to therapy. Moreover, some key clinical informative drivers were observed to show sex-biased clonal tendency in specific subtypes. As an extreme example, *TP53* mutation was reported as an early event in many cancers, including GBM.[27] However, in male patients with mesenchymal GBM, at least, this is not the case. According to our results, 100% of *TP53* mutations (8/8) were clonal in female patients with mesenchymal GBM, while only 46% of mutations (7/15) were observed to be clonal in male patients of the same subtype. This finding reminded us that sex and subtype should be considered simultaneously when developing new therapy strategies for glioma patients carrying specific alterations.

There were some limitations in this study. First, since TCGA predominantly concentrated on exome sequencing data, we were unable to investigate sex differences of mutation patterns in the noncoding regions. Second, our inference of mutation clonality was conducted within single tumor samples that were collected at one timepoint during the disease course, which may be an obstacle of deep analysis of tumor evolution. Therefore, multiregional sequencing and multi-timepoint sampling for a large patient cohort[8] are expected to comprehensively dissect sex differences during the cancer evolution.

IDH1/2 mutations have been proposed as early clonal events during gliomagenesis,[8,28] but in this study the clonal fraction of *IDH* mutation in LGG was observed to be only 53%. Using 2 independent datasets (Supplementary Methods), we demonstrated that the LGG samples in TCGA carry more subclonal *IDH* mutations than other data. The reason for this discrepancy is unclear. Single-cell mutation data of a large patient cohort are needed to further explore the clonality of *IDH* mutations in the future.

In conclusion, our results present a systematic characterization of mutation difference during glioma evolution between male and female patients and shed new insights, at the molecular level, on sex disparity underlying gliomagenesis. Finally, we expect that our analyses will be helpful for achieving the goal of precision medicine for brain tumor.

Keywords

cancer evolution | clonal status | diffuse gliomas | mutation | sex difference

Funding

This work was supported in part by the National High Technology Research and Development Program of China (863 Program, grant no. 2014AA021102), the National Program on Key Basic Research Project (973 Program, grant no. 2014CB910504), the National Natural Science Foundation of China (grant nos. 91439117, 61473106, and 61573122), the China Postdoctoral Science Foundation (grant no. 2016M600260), the Wu lien-teh youth science fund project of Harbin Medical University (grant no. WLD-QN1407), special funds for the construction of higher education in Heilongjiang Province (grant no. UNPYSCT-2016049), and the Heilongjiang Postdoctoral Foundation (grant no. LBH-Z16098).

Conflict of interest statement. The authors declare no conflicts of interest.

Authorship statement. HZ conceived this work, performed statistical analysis, and wrote the paper. JL, XZ, and EZ collected and pre-processed data, and wrote the paper with HZ. XL and SL helped to interpret the results. JS and FY generated the figures of subtype analyses. JX and WS helped to prepare the data of TCGA and other resources. YL helped to conceive this work. YX and XL supervised the whole study and revised the manuscript.

References

1. Schwartzbaum JA, Fisher JL, Aldape KD, Wrensch M. Epidemiology and molecular pathology of glioma. *Nat Clin Pract Neurol.* 2006;2(9):494–503; quiz 491 p following 516.
2. Dubrow R, Darefsky AS. Demographic variation in incidence of adult glioma by subtype, United States, 1992–2007. *BMC Cancer.* 2011;11:325.
3. Armstrong TS, Cao Y, Scheurer ME, et al. Risk analysis of severe myelotoxicity with temozolomide: the effects of clinical and genetic factors. *Neuro Oncol.* 2009;11(6):825–832.
4. Sun T, Warrington NM, Luo J, et al. Sexually dimorphic RB inactivation underlies mesenchymal glioblastoma prevalence in males. *J Clin Invest.* 2014;124(9):4123–4133.
5. Warrington NM, Sun T, Luo J, et al. The cyclic AMP pathway is a sex-specific modifier of glioma risk in type I neurofibromatosis patients. *Cancer Res.* 2015;75(1):16–21.
6. Sun T, Plutynski A, Ward S, Rubin JB. An integrative view on sex differences in brain tumors. *Cell Mol Life Sci.* 2015;72(17):3323–3342.
7. Wang J, Cazzato E, Ladewig E, et al. Clonal evolution of glioblastoma under therapy. *Nat Genet.* 2016;48(7):768–776.
8. Suzuki H, Aoki K, Chiba K, et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet.* 2015;47(5):458–468.
9. Brennan CW, Verhaak RG, McKenna A, et al; TCGA Research Network. The somatic genomic landscape of glioblastoma. *Cell.* 2013;155(2):462–477.
10. Brat DJ, Verhaak RG, Aldape KD, et al; Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N Engl J Med.* 2015;372(26):2481–2498.
11. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med.* 2015;7(283):283ra254.
12. Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell.* 2013;152(4):714–726.
13. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012;30(5):413–421.
14. Yuan Y, Liu L, Chen H, et al. Comprehensive characterization of molecular differences in cancer between male and female patients. *Cancer Cell.* 2016;29(5):711–722.
15. Kim H, Zheng S, Amini SS, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* 2015;25(3):316–327.
16. Nik-Zainal S, Van Loo P, Wedge DC, et al; Breast Cancer Working Group of the International Cancer Genome Consortium. The life history of 21 breast cancers. *Cell.* 2012;149(5):994–1007.
17. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 2016;131(6):803–820.
18. Marusyk A, DeGregori J. Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. *Biochim Biophys Acta.* 2008;1785(1):1–11.
19. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat.* 1985;39(1):33–38.
20. Verhaak RG, Hoadley KA, Purdom E, et al; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* 2010;17(1):98–110.
21. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell.* 2017;168(4):613–628.
22. Santarpia L, Lippman SM, El-Naggar AK. Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin Ther Targets.* 2012;16(1):103–119.
23. Meibohm B, Beierle I, Derendorf H. How important are gender differences in pharmacokinetics? *Clin Pharmacokinet.* 2002;41(5):329–342.
24. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer.* 2006;6(4):281–291.
25. Sun T, Warrington NM, Rubin JB. Why does Jack, and not Jill, break his crown? Sex disparity in brain tumors. *Biol Sex Differ.* 2012;3:3.
26. Jäger N, Schlesner M, Jones DT, et al. Hypermutation of the inactive X chromosome is a frequent event in cancer. *Cell.* 2013;155(3):567–581.
27. Attolini CS, Cheng YK, Beroukhim R, et al. A mathematical framework to determine the temporal sequence of somatic genetic events in cancer. *Proc Natl Acad Sci U S A.* 2010;107(41):17604–17609.
28. Johnson BE, Mazor T, Hong C, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science.* 2014;343(6167):189–193.