This is a repository copy of *Genomic landscape of diffuse glioma revealed by whole genome sequencing*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/226930/

Version: Published Version

Article:

Genomic landscape of diffuse glioma revealed by whole genome sequencing

Ben Kinnersley [1,2,30] ✉, Josephine Jung [3,4,30] ✉, Alex J. Cornish [1,30], Daniel Chubb[1,30], Ross Laxton[5], Anna Frangou[6], Andreas J. Gruber [7], Amit Sud [1], Giulio Caravagna[8], Andrea Sottoriva[8], David C. Wedge [9], Thomas Booth [10,11], Safa Al-Sarraj[5], Samuel E. D. Lawrence [1], Erminia Albanese[12], Giulio Anichini [13], David Baxter[14], Alexandros Boukas[15], Yasir A. Chowdhury[16], Pietro D'Urso[17], Robert Corns[18], Andrew Dapaah[19], Ellie Edlmann[20], Fay Greenway[21], Paul Grundy[22], Ciaran S. Hill [2,23], Michael D. Jenkinson[24], Sandhya Trichinopoly Krishna[24], Stuart Smith [19], Susruta Manivannan[22], Andrew J. Martin[21], Samir Matloob[25], Soumya Mukherjee[26], Kevin O'Neill[13], Puneet Plaha [15], Jonathan Pollock[25], Stephen Price [26], Ola Rominiyi [27], Bobby Sachdev [12], Fozia Saeed[18], Saurabh Sinha[27], Lewis Thorne [23], Ismail Ughratdar[16], Peter Whitfield[20], Amir Saam Youshani [17], Helen Bulbeck [28], Prabhu Arumugam[29], Richard Houlston [1,31] ✉ & Keyoumars Ashkan[3,4,31] ✉

Diffuse gliomas are the commonest malignant primary brain tumour in adults. Herein, we present analysis of the genomic landscape of adult glioma, by whole genome sequencing of 403 tumours (256 glioblastoma, 89 astrocytoma, 58 oligodendroglioma; 338 primary, 65 recurrence). We identify an extended catalogue of recurrent coding and non-coding genetic mutations that represents a source for future studies and provides a high-resolution map of structural variants, copy number changes and global genome features including telomere length, mutational signatures and extrachromosomal DNA. Finally, we relate these to clinical outcome. As well as identifying drug targets for treatment of glioma our findings offer the prospect of improving treatment allocation with established targeted therapies.

Diffuse gliomas are the most common malignant primary brain tumour in adults[1,2]. Due to their infiltrative nature gliomas typically have a very poor prognosis irrespective of clinical care with the most common type, glioblastoma (GBM) having a median survival of only 12 to 15 months[3].

Traditionally, classification and prognostication of gliomas has been solely based on tumour morphology. Recently molecular characterization has yielded more uniform disease entities that better capture biological and clinicopathological features[4–12]. This has informed version five of the WHO classification with molecular features dictating subtype and joint histopathologic and molecular analysis determining grade resulting in three primary tumour groups: IDH-mutant, 1p/19q co-deleted oligodendroglioma; IDH-mutant, non-co-deleted astrocytoma; and IDH-wildtype GBM[13].

There is increasing evidence that the clinical behaviour of tumours and their therapeutic response is influenced by tumour genomics. However, thus far sequencing of glioma has largely been confined to protein-coding (exome) regions, and a comprehensive description of its genomic landscape is lacking. Specifically, there is a paucity of information on focal copy number

changes, structural variants, and mutations within non-coding elements[14–18].

In this work we detail the genomic landscape of glioma by WHO v5 subtype through analysis of whole-genome sequencing (WGS) data on 403 patients recruited to the 100,000 Genomes Project (100kGP). The results of our study provide additional insights into coding and non-coding single nucleotide mutations. We also provide a detailed map of structural alterations and global features, including telomere length, mutational signatures, chromothripsis and extrachromosomal DNA. Finally, we relate these genomic features to clinical outcome. As well as identifying drug targets our findings offer the prospect of improving treatment allocation with targeted therapies.

Results

The 100 kGP cohort

The analysed cohort (100 kGP, release v8) comprised tumour-normal (T/N) sample pairs recruited to 100 kGP through 13 NHS Genomic Medicine Centres across England. The histology of each patient's tumour was reviewed by a neuropathologist and glioma subtype assigned according to WHO v5. Comprehensive clinicopathological information on patients is provided in Supplementary Data 1–2 and Fig. 1. We restricted our WGS analysis to high-quality data from flash-frozen tumour samples (Online Methods). In addition to using variant calls from the 100 kGP analysis pipeline we: (i) removed alignment bias introduced by ISAAC soft clipping of semi-aligned reads[19]; (ii) called tumour copy number using Battenberg[20]; (iii) called structural variants (SVs) from the consensus of Manta[21], LUMPY[22], and DELLY[23]; (iv) removed indels within 10 base pairs (bp) of a common germline indel. Complete details on sample curation, somatic variant calling, and annotation of mutations are provided in Online Methods. After quality control sequenced cancers comprised 58 IDH-mutant, 1p/19q co-deleted oligodendroglioma (19 recurrent), 89 IDH-mutant, non-co-deleted astrocytoma (22 recurrent) and 256 IDH-wildtype GBM (24 recurrent) (Fig. 1).

Ten of the patients (2.5%) were carriers of pathogenic germline variants in one of the genes previously reported to influence glioma susceptibility (POLE n = 2, MLH1 n = 1) or one of other well-established cancer susceptibility genes which play a role in DNA repair (ATM n = 2, BRCA1 n = 1, BRCA2 n = 2, OGG1 n = 2) (Supplementary Data 3).

In keeping with WHO v5, ATRX (72%) and TP53 (97%) mutations were common in astrocytoma and TERT promoter (TERTp) mutations common in GBM (91%). As expected, higher tumour grade was associated with a significantly worse overall survival (OS) and progression-free survival (PFS) in astrocytoma and oligodendroglioma. Similarly, MGMT methylation status was predictive of OS and PFS in GBM (Fig. 1, Supplementary Data 1).

Across all T/N pairs, as per previous reports[4], mutational burden varied significantly between glioma subtypes with IDH-wildtype GBM having a significantly higher rate of SNVs, indels, copy number alterations and structural variants (Supplementary Data 4). Ten tumours (7 GBMs, 1 astrocytoma and 2 oligodendrogliomas) displayed a hypermutated phenotype (i.e., SNV/indel mutation burden >10 mutations per Mb).

The global pattern of somatic alterations in the 403 tumours is shown in Fig. 2. Each of the glioma subtypes were associated with multiple chromosomal arm level changes, the commonest being gain of chr7 (GBM 84%, astrocytoma 23%) and chr11 (oligodendroglioma 10%), and deletion of chr10 (GBM 84%), chr9p (astrocytoma 26%) and chr4 (oligodendroglioma 17%) (Supplementary Data 5; Fig. 2).

Driver mutations

Protein-coding driver gene identification was performed using IntO-Gen, which incorporates seven complementary algorithms[24]. Across all glioma subtypes, 42 genes were significantly somatically mutated, comprising 29 established glioma drivers, seven established in other cancer types (BCORL1, CREBBP, PBRM1, PPM1D, SETD1B, TRRAP, VAV1) and six (CDC73, COL3A1, SALL3, SFPQ, TENM2, TMEM132C), which either had not been reported previously or have had frequencies of less than 1% in landmark studies (Supplementary Data 6–8).

The known glioma drivers within tumour subgroups were mutated at close to reported frequencies (e.g. ATRX and TP53 mutations in astrocytoma, CIC, NOTCH1 and FUBP1 in oligodendroglioma, EGFR, PTEN, RB1 in GBM, IDH1 and IDH2 with astrocytoma and oligodendroglioma, and TERTp mutations with GBM and oligodendroglioma; Supplementary Data 6–7). Given the scale of previous exome sequencing studies, as expected all the novel coding drivers were at low frequencies (1.2–4.5%; Supplementary Data 6–7; Supplementary Fig. 1).

As expected from previously published work[6,13,25], each subtype was associated with a core set of mutated driver genes (carrier frequency >5%): GBM by TERT, PTEN, EGFR, TP53, NF1, RB1, PIK3CA and PIK3RI mutations; astrocytoma by IDH1, ATRX, TP53 and PIK3CA mutations; and oligodendroglioma by TERT, IDH1, CIC, FUBP1, PIK3R1, NOTCH1, ZBTB20, ARID1A, BCOR, IDH2 and SETD2 mutations (Fig. 2).

Candidate driver genes generally had low numbers of mutations that could confidently be classified as oncogenic, with the majority classified as variants of uncertain significance (VUS). Clonal frequencies of mutations in driver candidates ranged from 54% to 86%, consistent with these driver events occurring later than those in established drivers (Supplementary Data 9). We assessed the frequency of copy-number gains or LOH events at candidate drivers for predicted oncogenes or TSGs, respectively, notably finding LOH events in 15% (62/403) and 8.4% (34/403) of tumours at SFPQ (chr1p34.3) and SALL3 (chr18q23), respectively (Supplementary Data 10). However, we could not detect any "second hit" events at candidate drivers, likely due to rarity of their mutation. We surveyed information from DepMap gene perturbation screening[26], which provided further relevance for SFPQ in the biology of glioma (Supplementary Data 11).

While acknowledging the absence of available supporting transcriptomic data, biological mechanisms highlighted by the candidate drivers included transcriptional and post transcriptional control pathways (CDC73[27], CREBBP[28]), ADGRG1-associated neuronal migration and RhoA-pathways (COL3A1[29], VAV1[30]). Additionally, the new candidate drivers highlight the role of epigenetic modification in the development of glioma through SWI/SNF mediated chromatin remodelling (PBRM1[31]) and the SET1 histone methyltransferase complex (SETD1B[32]).

We searched for non-coding drivers using OncodriveFML[33] to detect positive selection across promoters, enhancers, non-canonical splice regions, untranslated regions (UTRs), microRNAs (miRNAs), long intergenic non-coding RNAs (lincRNAs), CCCTC-binding factor (CTCF) sites, transcription factor (TF)-binding sites, and regions of open chromatin. As expected, TERTp mutations predominated, being detected in 91%, 1% and 100% of GBM, astrocytoma and oligodendroglioma tumours, respectively (Supplementary Fig. 2; Supplementary Data 12). Aside from the "canonical" C228T and C250T mutations (with both found in one GBM tumour), four other TERTp variants were identified in oligodendrogliomas. Intriguingly, while one variant (chr5:1295208G>A) co-occurred with C250T, the other variants were present in the three oligodendrogliomas lacking a canonical TERTp mutation (chr5:1295061G>C & chr5:1295066G>A, chr5:1295046T>G, chr5:1295060C>CGAGGAGAGGGCGGGGCCGCGGAAAGGAAGGGGAGGGGCTGG); Supplementary Data 13 and 14).

In addition to TERTp mutations we identified variants in nine other recurrently mutated non-coding elements (Supplementary Fig. 2; Supplementary Data 12). These included, the 3'UTR of CCDC85 at 14q32.3 (3% of gliomas), the 7p12.2 enhancer near VWC2 (2%), the CTCF site within CROCC at 1p36.13 (2%), the open chromatin region at 15q11.2 (1.5%), as well as the brain GeneHancer[34] elements within EGFR (7p11.2;
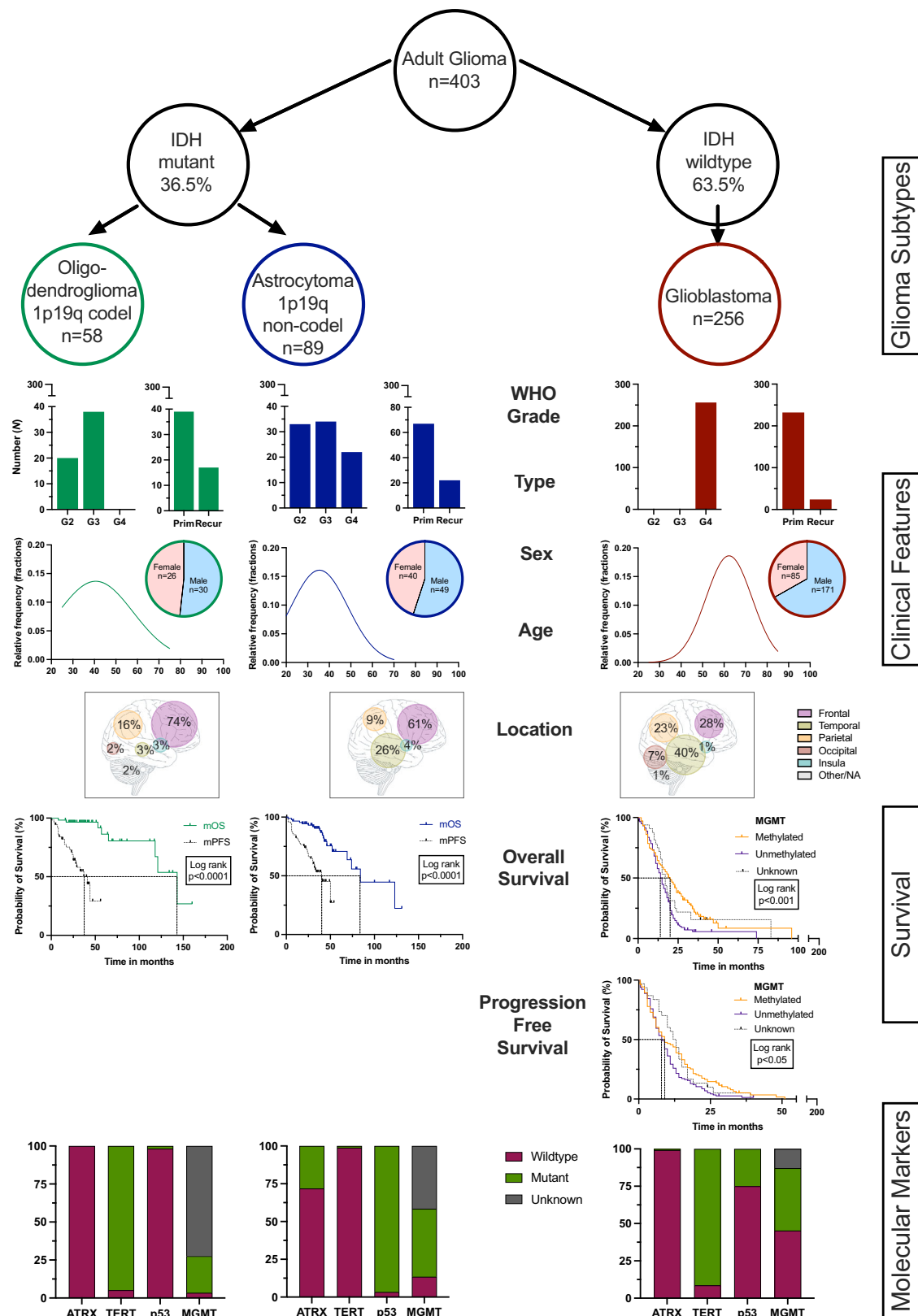
4%), near *VWC2* (7p12.2, 2%), *ITGA1* (5q11, 3%), *LINC01787* (1p21.1, 3%) and *FEZ2* (2p22.2, 6%). While a role for mutation of these non-coding elements in glioma oncogenesis is yet to be established, two have speculative biological plausibility a priori: *CCDC85C* (coiled-coil domain containing 85c) is implicated in neurogenesis, gliogenesis, and ependymogenesis[35], and upregulation of *ITGA1* (Integrin Subunit Alpha 1) is implicated in temozolomide (TMZ) resistance[36]. Additionally, mutations in the GeneHancer element within *EGFR* (chr7:55107542-55114199) were associated with adverse survival in patients with recurrent oligodendroglioma ($P_{\text{logrank}} = 7.5 \times 10^{-5}$; Supplementary Fig. 3; $P_{\text{coxph}} = 0.02$; OR = 31.66, 95% confidence interval (CI) 1.63–616.10; Supplementary Data 15). Variants within this region are

Fig. 1 | Study overview. Overview of clinical features (WHO grade 2–4, primary vs recurrent tumour–top row; gender and age distribution–second row) stratified by tumour subtype. Left column (green) depicts oligodendroglioma (*n* = 58), middle column (blue) depicts astrocytoma (*n* = 89) and right column (red) depicts glioblastoma (*n* = 256). Schematic representation of tumour location for each glioma subtype by lobe (frontal, temporal, parietal, occipital, insula, other location) in the third row. Survival is shown as overall survival and progression free survival classified by tumour grade for oligodendroglioma and astrocytoma (overall survival–coloured solid line, progression free survival–black dashed line). Overall survival and progression free survival for glioblastoma is shown by MGMT methylation status (methylated–coloured solid line, unmethylated–coloured dashed line, unknown–black line). Kaplan-Meier curves are plotted with significance between groupings estimated using the log-rank test. *P*-values are two-sided and unadjusted. Last row demonstrates molecular markers stratified by tumour subtype (wild type–pink, mutant–green, unknown–grey).
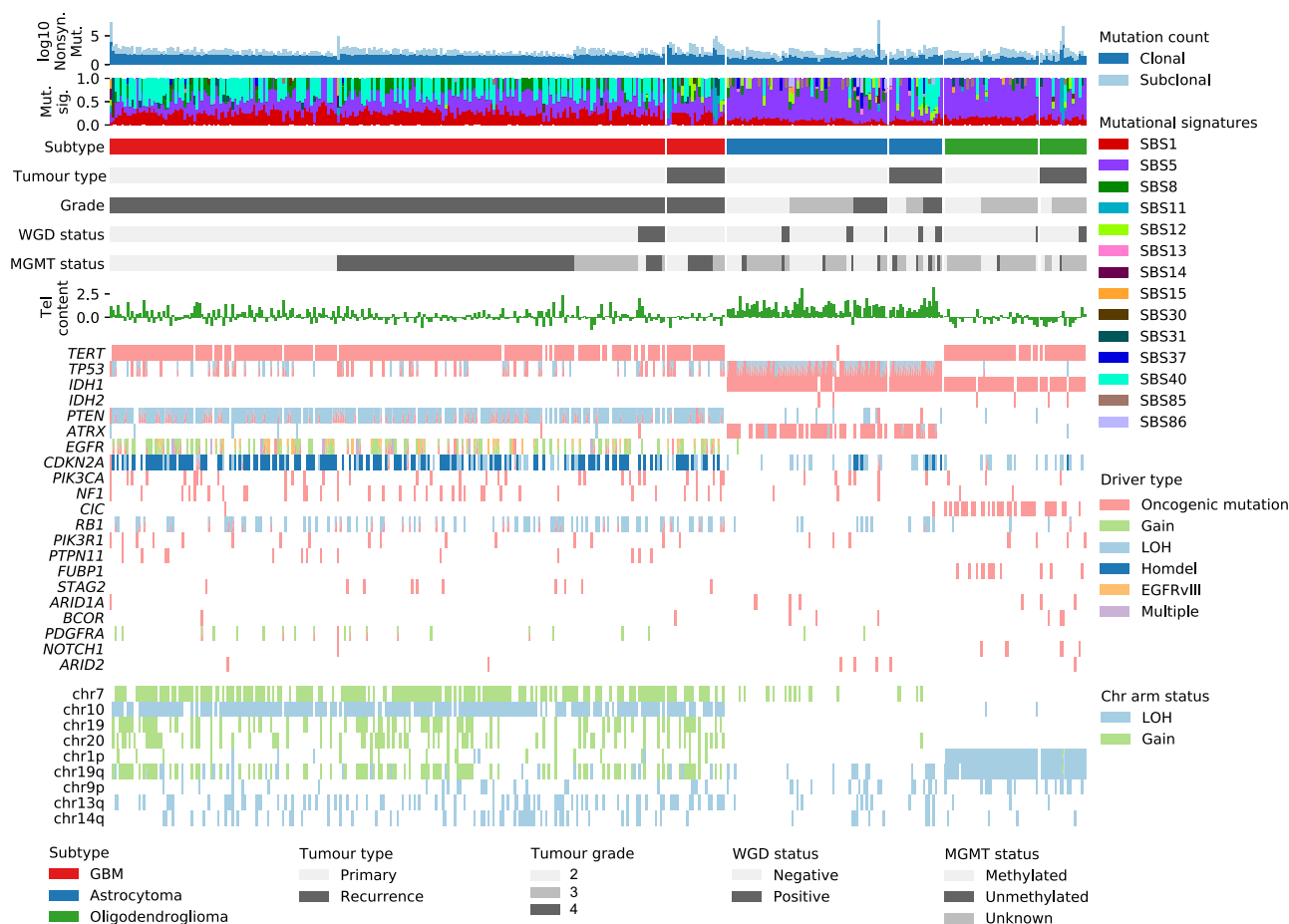

Fig. 2 | Summary of molecular alterations across glioma subtypes. From top to bottom: clonal and subclonal nonsynonymous mutation counts (log10); proportions of single base substitution (SBS) mutational signatures; glioma subtype; tumour type; tumour grade; whole-genome duplication (WGD) status; MGMT promoter methylation status; normalised log2(tumour/normal) telomere content; alteration status of common glioma driver genes; gain or loss of heterozygosity (LOH) status of chromosome arms commonly gained or lost in glioma. Homdel, homozygous deletion.

associated with *EGFR* expression in the Genotype-Tissue Expression (GTEx) portal[37], consistent with the region being a cis-regulatory element for *EGFR*. However, mutations in chr7:55107542-55114199 showed no association with either oncogenic *EGFR* mutations or copy number gains (Supplementary Data 16).

Recurrent structural variants and copy-number alterations
We used GISTIC2[38] to identify genomic regions recurrently affected by focal amplifications and deletions and infer likely target genes (Q < 0.05, frequency >10%; Supplementary Fig. 4, Supplementary Data 17). In GBM copy number alterations (CNAs) implicated *EGFR* (amp7p11.2), *CDK4* (amp12q14.1), *PDGFRA* (amp4q12), *MDM4* (amp1q32.1), *SOX2* (amp3q26.33), *CDKN2A* (del9p21.3), *QKI* (del6q26), *PTEN* (del10q23.31), *B2M* (del15q14), *RB1* (del13q14.2). In astrocytoma, *MYC* (amp8q24.21), *CCND2* (amp12p13.32), *FAT1* (del4q35.2), *CDKN2A*

(del9p21.3) and *MGMT* (del10q26.3) were highlighted as driver genes. We further identified 59 focal events (8 GBM, 9 astrocytoma, 1 oligodendroglioma) for which no driver gene could be unambiguously ascribed, of which 8 (2 GBM (amp7q11.23, amp19p13.3), 6 astrocytoma (amp3p26.3, amp10p15.3, amp11q23.3, amp11q25, amp12q24.33, del5p15.33)) had not been previously been reported[39]. Intriguingly, the astrocytoma del5p15.33 CNA encompasses *TERT*, and amp11q23.3 overlaps *PHLDB1* which is implicated in genetic susceptibility to non-GBM glioma[40]; however, whether these are the target genes of the focal CNAs is yet to be determined (Supplementary Data 17).

We identified 22 hotspots of recurrent simple SVs (Q < 0.05) by piece-wise constant fitting[41], adjusting for local genomic factors known to influence rearrangement density (i.e. chromatin accessibility, repeated elements, GC content, replication timing, gene density and expression)[42] (Q < 0.05, Supplementary Data 18). Fragile sites are
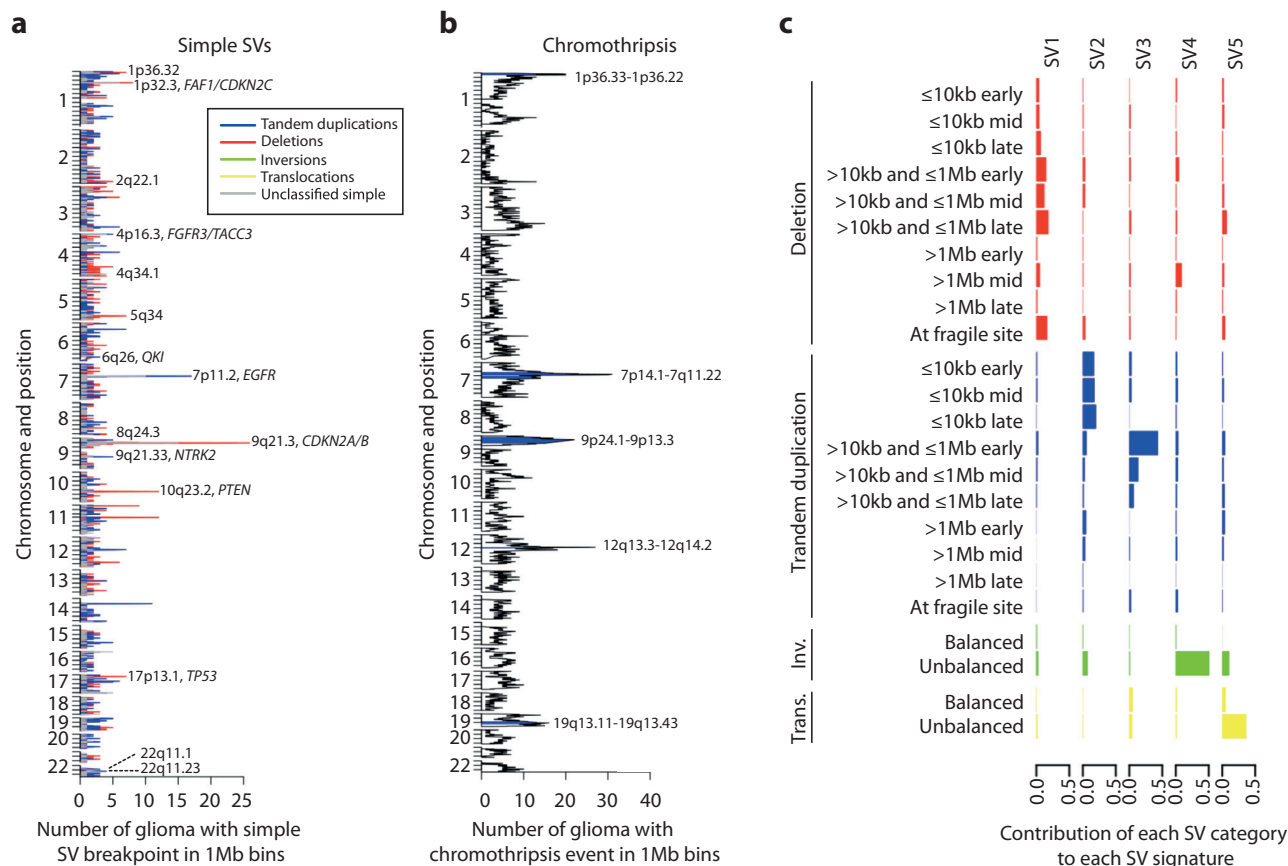
Fig. 3 | Patterns of structural variation. a Significant hotspots of simple structural variants (SVs) identified in all unique glioma samples (*n* = 400). Non-fragile SV hotspots identified at a false discovery rate (FDR) of 5% are annotated with the cytoband and any candidate genes. Fragile site SVs have not been plotted. Coloured lines represent the number of tumours with a SV break point of each type in 1 Mb genome regions. **b** Frequency of chromothripsis events, with regions enriched for chromothripsis at a 5% FDR greater than 5 Mb are coloured blue. **c** 5 SV signatures extracted from glioma samples with ≥10 simple SVs. Bars represent each SV category's contribution to each SV signature. Inv.: inversions; Trans.: translocations.

prone to rearrangement (possibly due to replication error), tending to co-occur with large and late-replicating genes[43], hence SVs occurring at such sites are the consequence of mechanistic rather than selective factors[44]. In view of this we excluded seven SV hotspots mapping to potential fragile sites from further analysis (Supplementary Data 18).

Seven of the 15 SV hotspots we identified occurred at sites previously reported as being either recurrently deleted (del9p21.3 - *CDKN2A/B*, del10q23.2 - *PTEN*, del6q26 - *QKI*, del17p13.1 - *TP53*) or amplified (amp7p11.2 - *EGFR* and amp9q22.3 - *NTRK2*) in addition to fusions (4p16.3 - *FGFR3-TACC3*)[4,45] (Fig. 3a; Supplementary Data 18; Supplementary Fig. 5). Since 1p32.3 deletions spanned both *FAF1* and *CDKN2C*[46,47] it is unclear whether a functional consequence of this rearrangement is through dysregulation of one or both genes. The del5q34 implicates the long non-coding RNA LOC102546299 in glioma, while the target gene of amp22q11.23 remains to be established (Supplementary Data 18; Supplementary Fig. 5). A number of SV hotspots were associated with OS in primary GBM, with favourable prognosis at amp8q24.3 (P_{coxph} = 0.003; HR = 0.26, 95% CI 0.10−0.63) and worse prognosis at del10q23.2 (P_{coxph} = 0.018; HR = 1.74, 95% CI 1.10−2.75), del6q26 (P_{coxph} = 0.035; HR = 1.42, 95% CI 1.02−1.97) and del2q37.2 (P_{coxph} = 0.04; HR = 1.46, 95% CI 1.02−2.11; Fig. 4; Supplementary Data 15; Supplementary Fig. 3).

To investigate the mutational processes underlying SV formation, we first grouped simple SVs into 24 categories, considering multimodal size distributions, replication timing and whether deletions and

tandem duplications were located at fragile sites. Application of a hierarchical Dirichlet process (HDP)[18] extracted five SV signatures (referred herein as SV1-5; Fig. 3c). As per previous analyses[18], we extracted signatures comprising small and medium-sized deletions (SV1), small tandem duplications (SV2), medium tandem duplications (SV3), inversions (SV4) and translocations (SV5).

We examined for associations between extracted SV signatures and SVs occurring at recurrent hotspots (P < 0.05; Supplementary Data 19). In primary GBM, SV1 activity was significantly higher at the hotspots del1p32.3 (*FAF1/CDKN2C*), del4q34.1, del10q23.2 (*PTEN*) and del22q11.1. In recurrent GBM, SV5 was associated with del10q23.2 (*PTEN*), and in primary astrocytoma, SV4 activity was significantly lower at del2q37.2 (Supplementary Data 19).

Across all tumours, chromothripsis was detected on chromosomes 1, 7, 9, 12 and 19 (Q < 0.05; Fig. 3b). This was more common in GBM (73%) and astrocytoma (55%) compared with oligodendroglioma (3%; Supplementary Fig. 6). Several of the chromothripsis-enriched regions are sites of recurrent gains or deletions containing tumour suppressor or oncogenes−7p14.1-7q11.22 (*EGFR*) 12q13.3-12q14.2 (*CDK4*)[48], and 9p24.1-9p13.2 (*CDKN2A/B*)[4]. Whether the chromothripsis events at 1p36.33-1p36.22 and 19q13.11-19q13.43 are a consequence of a yet unidentified selective process remains to be established. We identified very few chromoplexy events (Supplementary Fig. 7) while several regions were enriched for complex unclassified SV events (Supplementary Fig. 8).
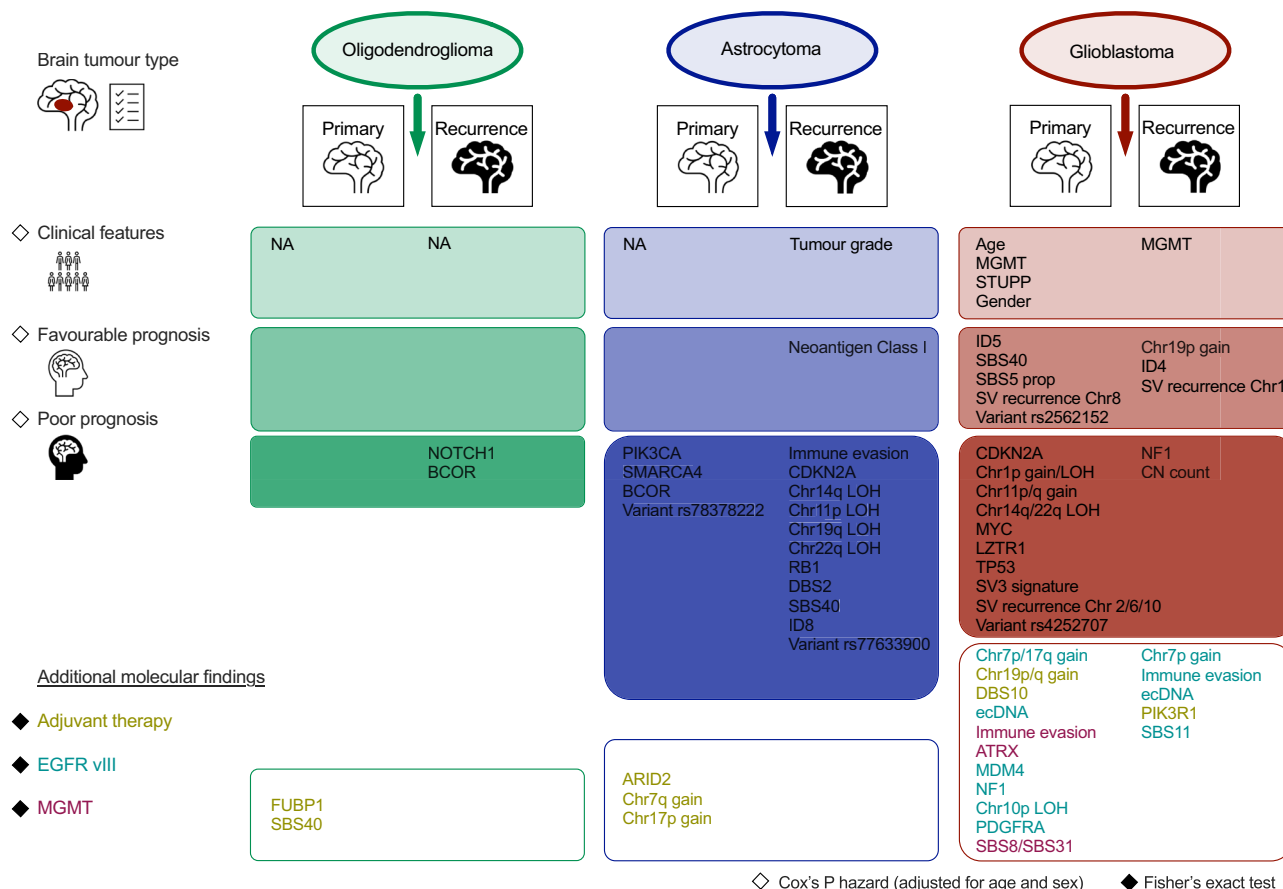
Fig. 4 | Prognostic insights depicted per tumour subtype. Genetic alterations found in Oligodendrogliomas (green), Astrocytomas (blue) and Glioblastoma (red) that are associated with a favourable (middle row) or unfavourable prognosis (bottom row). Additional molecular findings associated with adjuvant therapy, EGFR vIII mutation and MGMT status are shown at the bottom.

Extrachromosomal DNA

The WGS data allowed us the explore the landscape of extrachromosomal DNA (ecDNA) formation in glioma. Amplicon structures were detected and classified using Amplicon Architect[49]. As per previous reports[50], ecDNA was primarily a feature of GBM (54%; 137/256) rather than astrocytoma (7%, 6/89) or oligodendroglioma (2%, 1/58) (Supplementary Fig. 9, Supplementary Data 20). There was no significant difference in the frequency of ecDNA between primary (54%, 126/232) and recurrent (46%, 11/24) GBMs (P_{fisher} = 0.52; Supplementary Data 16 and 20). As previously documented[51], GBMs with ecDNA were more likely to exhibit chromothripsis (P_{Fisher} = 7.6 × 10^{-3}; OR = 2.2; Supplementary Data 20). In addition to *EGFR* (64%, 152/237), ecDNA-mediated oncogenic amplification of *CDK4* (9%), *MDM2* (5%), *MDM4* (5%) and *PDGFRA* (3%) was a feature of GBM. ecDNA-mediated oncogenic amplification has been purported to be associated with poor clinical outcome in many cancers[51]. While ecDNA-mediated *EGFR* amplification was associated with significantly shorter OS when considering all gliomas (Supplementary Fig. 10a; P_{LRT} < 0.0001; median OS 16 vs. 25 months), when restricted to primary GBM there was no statistically significant difference (P_{LRT} = 0.39; median OS 17 vs. 15 months; Supplementary Fig. 10b). This was also the case stratifying by *MGMT* methylation status in GBM (Supplementary Fig. 10d). It has also been suggested that ecDNA leading to *EGFR* amplification in GBM mediates resistance to EGFR inhibitors[52] and that ecDNA is a basis of treatment-resistance[53]. However, in our analysis there was no evidence to support such an assertion in tumours treated with the "standard" regimen (i.e. radiotherapy and temozolomide).

After comparing the total genomic fraction altered by copy number (CN) and total CN, we classified tumours into four CN profiles: 1, representing "quiet CN" tumours; 2, representing "broad CN" tumours dominated by arm-level events; 3, representing "focal CN" tumours dominated by focal events; 4, representing "disrupted CN" tumours with large numbers of arm-level and focal changes (Supplementary Fig. 11). As expected, the vast majority of oligodendrogliomas were CN quiet (primary [P]: 85%, recurrence [R]: 74%), with the proportion of tumours dominated by broad CN changes increasing with recurrence (P: 15%, R: 26%). Primary astrocytomas were similarly CN quiet (P: 72%, R: 41%) while at recurrence there was an increase in disrupted CN tumours (P: 12%, R: 50%). In contrast for GBM the minority of tumours were CN quiet (P: 16%, R: 8%) while the dominant category was CN disrupted tumours (P: 47%, R: 58%; Supplementary Data 21). We compared driver mutation frequencies across these CN profiles and in primary GBM tumours found a subset of drivers enriched for CN quiet tumours: *CDKN2A* (P_{fisher} = 0.031), *PDGFRA* (P_{fisher} = 0.034), *PPM1D* (P_{fisher} = 0.025), and another subset enriched for CN broad tumours: *PIK3CA* (P_{fisher} = 0.01), *PTEN* (P_{fisher} = 0.026), *RB1* (P_{fisher} = 0.038; Supplementary Data 21).

Mutational signatures

To gain insight into mutational processes in glioma, we extracted single-base substitution (SBS), double-base-substitution (DBS) and insertion-deletion (indel; ID) signatures de novo and related them to known COSMIC signatures (v3.2) using SigProfilerExtractor[16,54] (Fig. 5; Supplementary Figs. 12–16; Supplementary Data 22). In most gliomas,
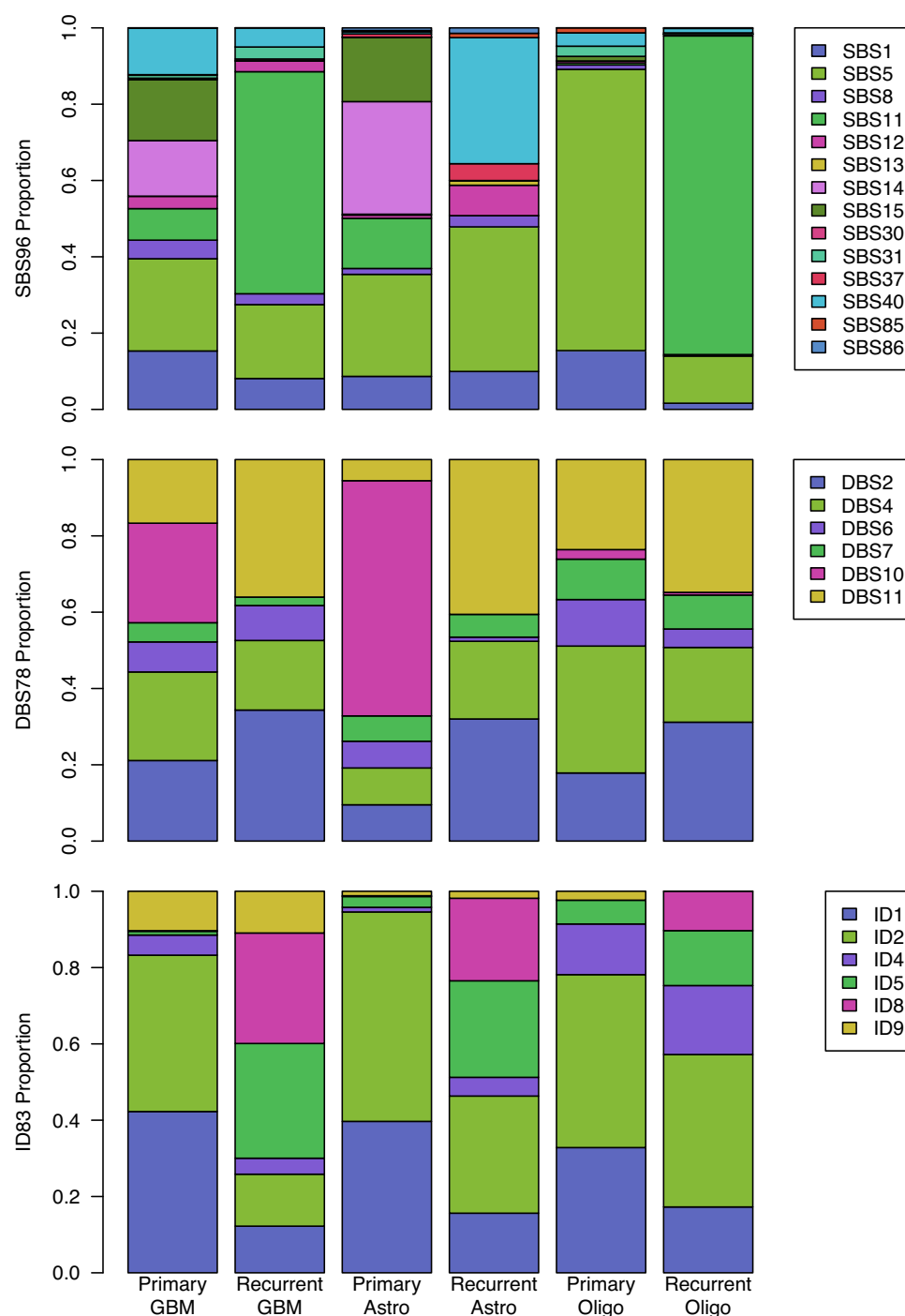
Fig. 5 | Mutational signatures across subtypes. Plotted are overall proportions of SBS96, DBS78 and ID83 signature activities detected across each glioma subtype.

the major SBS signatures changes were SBS5/SBS40 and SBS1 resulting from clock-like mutagenic processes (Fig. 5).

While SBS signatures extracted were broadly consistent with those reported previously for low-grade glioma and GBM[16], some tumours exhibited more unusual and specific SBS patterns. Notably a signature attributable to combined DNA polymerase epsilon (*POLE*) exonuclease domain mutations and mismatch repair deficiency (SBS14) was detected in three tumours (two GBM, one grade 4 astrocytoma). These patients were carriers of either pathogenic germline mutations in *POLE* (*n* = 2) or *MLH1* (*n* = 1). Both *POLE* carriers had a previous history of cancer. Tumours from these three patients did not have *TERTp* mutation, EGFR amplification, chr7 gain or chr10 loss (Supplementary Data 23). These features are consistent with the

previously described phenotype associated with "ultramutated" high-grade glioma associated with somatic *POLE* mutation[55], specifically bizarre multinucleated giant cells and eccentric nuclei (Supplementary Data 23). There was, however, variability in the ratio of clonal to subclonal mutations between the three tumours, suggestive of differences in the timing of acquisition of mismatch repair deficiency and *POLE* mutation (Supplementary Fig. 17).

SBS18 and SBS30 have both been postulated to result from damage by reactive oxygen species (ROS)[56]. SBS30 was detected in 15 tumours (13 GBM, 2 oligodendroglioma), which has been attributed to defective base excision repair, and specifically defective *NTHL1*[57]. Analyses of respective samples however provided no evidence to support inactivation of *NTHL1* through

Fig. 6 | Driver mutations attributed to SBS96 and ID83 mutational signatures. Oncogenic driver mutations were attributed to signature probabilities based on mutational contexts (Supplementary Methods).

germline or somatic mutation as a causal basis (Supplementary Data 3 and 16). A signature attributed to platinum chemotherapy exposure, SBS31, was detected in 27 tumours and as expected SBS11 which is associated with temozolomide treatment was highly enriched in recurrent tumours (P_{Fisher} = 4.57 × 10^{-5}; OR = 10.7; Fig. 5; Supplementary Data 16). While we did not identify any novel ID or DBS signatures, ID4 was recovered in 141 tumours (119 GBM, 10 astrocytoma, 12 oligodendroglioma). ID4 has recently been shown to be caused by defective ribonucleotide excision repair implicating topoisomerase 1 (TOP1) activity as a mechanistic basis[58]. This is of clinical relevance since irinotecan, which has been proposed as a therapy for glioma, targets TOP1[59]. ID8, which has been ascribed to DNA double-strand break repair by non-homologous end joining, was associated with tumour recurrence in all tumour subtypes (Fig. 5; Supplementary Data 16). Similarly, DBS11 is associated with recurrence in GBM tumours and is highly correlated with ID8 (Supplementary Data 16 and 24).

In general there was a trend toward a lower proportion of sub-clonal mutations attributable to the "clock-like" signatures SBS1, SBS5 and SBS40 in primary GBM tumours, and higher proportion of "treatment-associated" subclonal SBS11 and ID8 mutations in recurrent GBM tumours (Supplementary Fig. 18). Considering the signatures attributed to driver mutations, primary tumours are dominated as expected by the clock-like signatures SBS1, SBS5 and SBS40, with a small number of drivers in recurrent tumours attributed to treatment-related signatures, such as *RB1* and *SMARCA4* with SBS11 in recurrent GBM and *SETD2* and *CREBBP* with SBS11 in recurrent oligoden-droglioma (Fig. 6).

We examined for correlations between identified mutational and SV signature activities across all diffuse glioma tumours (Supplementary Fig. 19; Supplementary Data 24). SV1 (small and medium-sized deletions) and SV3 (medium tandem duplications) activities were negatively correlated, while SV3 exhibited significant positive correlation with SBS86 (attributed to unknown chemotherapy treatment[60]) and negative correlation with SBS31. SV3 and SV5 (translocations) were negatively correlated, while SV5 was significantly positively correlated with DBS6, ID2, ID4 and ID8. These findings are consistent with different SV classes being associated with distinct mutational processes.

Timing of driver events

We used MutationTimeR[17] to classify SNV and indel mutations as clonal or subclonal (Supplementary Fig. 20). There was a minor difference in the subclonal to clonal ratio between primary and recurrent GBM, astrocytoma or oligodendroglioma (Supplementary Data 4). In GBM the established drivers, *TERTp*, *PTEN*, *TP53*, *EGFR*, were essentially clonal, consistent with their early acquisition in tumour development (Fig. 7; Supplementary Data 9). In astrocytoma, while most *TP53*, *IDH1* and *ATRX* mutations were clonal, in contrast to GBM a high proportion, albeit at low overall frequency, of *NF1* mutations were subclonal, implying they are not initiating events. In oligodendrogliomas, as expected *IDH1/IDH2* mutations were almost exclusively clonal as well as most *TERTp* mutations. Most *CIC* and *FUBP1* mutations, however, were subclonal. Since these genes map to 19q13.2 and 1p31.1, it is likely that following the initiator 1p/19q co-deletion driver event the "second hit" occurs later in the evolution of these tumours (Fig. 4; Supplementary Data 9).
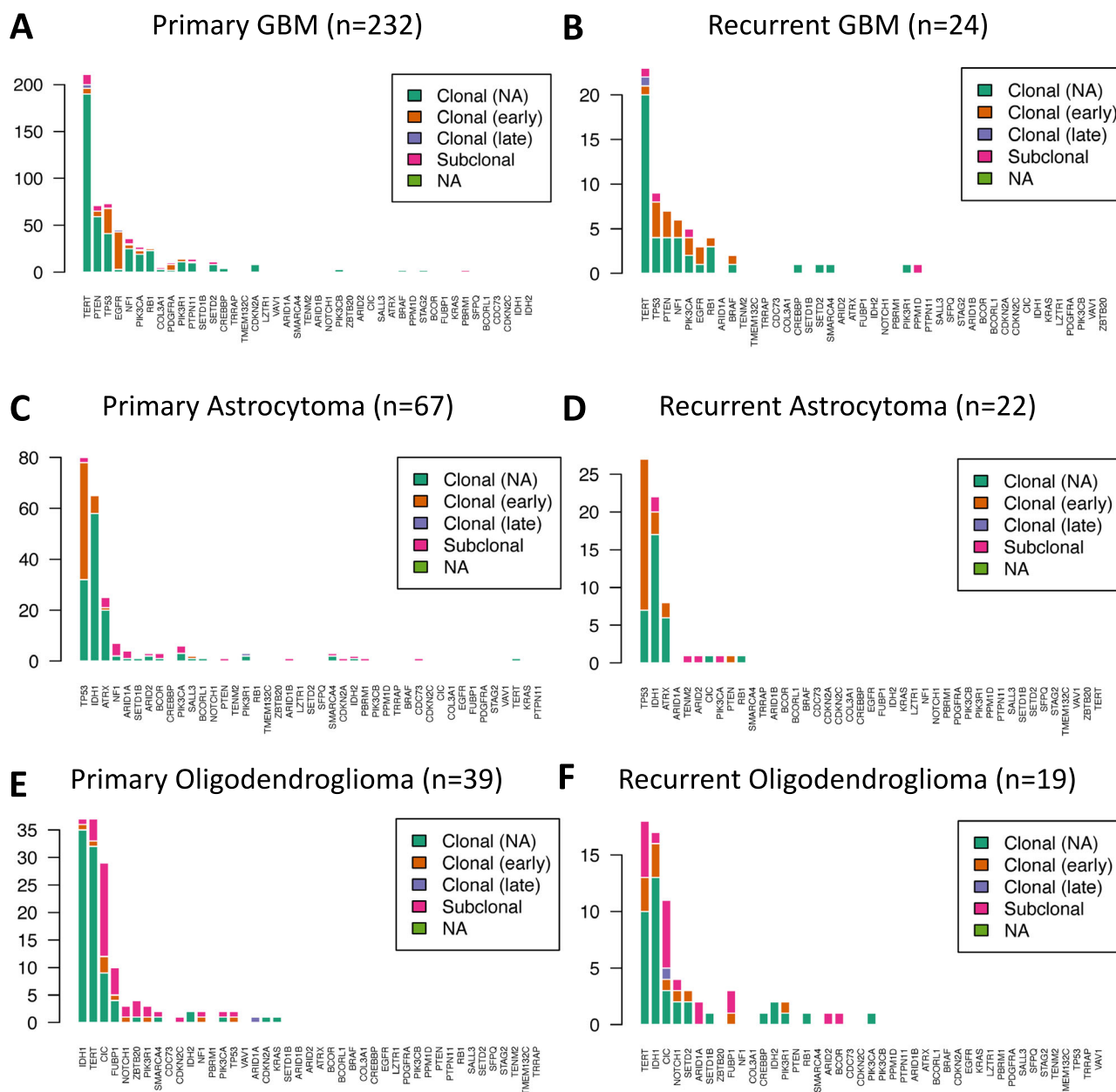
Fig. 7 | Driver mutation clonality by subtype. A Primary GBM ($n = 232$); **B** recurrent GBM ($n = 24$); **C** primary astrocytoma ($n = 67$); **D** recurrent astrocytoma ($n = 22$); **E** primary oligodendroglioma ($n = 39$); **F** recurrent oligodendroglioma ($n = 19$). Total $n = 403$.

Telomere maintenance

Telomerase dysfunction is a characteristic of glioma, with *TERTp* and *ATRX* mutations affecting "telomerase activation" (TA) and "alternative lengthening of telomeres" (ALT) pathways, respectively. To examine the relative contribution of TA- and ALT-pathway disruption to the development of tumour subtypes we estimated telomere content using Telomere Hunter (TH)[28] and Telomerecat (TC)[29]. Normalised tumour telomere content varied significantly between GBM, astrocytoma and oligodendroglioma tumours (Fig. 8A; Supplementary Fig. 21a; Supplementary Data 25). In primary GBM there was a significant difference between tumours without *TERTp* or *ATRX* mutations and those with *TERTp*, with the distribution of "neither" appearing like the ALT phenotype associated with ATRX-loss in astrocytoma (Fig. 8B; Supplementary Data 25). Intriguingly, primary GBMs without gain of chr7q or *TP53* mutations were characterised by longer telomeres (chr7q gain $P_{TC} = 4.81 \times 10^{-5}$; $P_{TH} = 1.32 \times 10^{-4}$ and *TP53* mutations $P_{TC} = 2.63 \times 10^{-3}$; $P_{TH} = 8.19 \times 10^{-5}$; Supplementary Data 4).

Immune evasion

Recurrent GBM tumours had significantly more total, clonal and subclonal mutations and neoantigens than primary GBMs, while there was no difference between primary and recurrent astrocytomas or oligodendroglimas (Supplementary Figs. 22 and 23). There was also a significant difference in the proportion of clonal and subclonal mutations and neoantigens between subtypes, with GBMs having significantly higher numbers of clonal mutations compared with astrocytomas and oligodendrogliomas ($P_{mut} = 9.58 \times 10^{-33}$; $P_{neo} = 1.96 \times 10^{-24}$; Supplementary Fig. 24). As expected, neoantigen and total mutation counts were highly correlated (Supplementary Figs. 25a and 26). In general, however there was poor correlation between clonal and subclonal neoantigen counts, and we observed a small number of hypermutated outliers in the primary GBM and primary astrocytoma subtypes—one exhibiting a large excess of subclonal compared to clonal neoantigens (Supplementary Fig. 25b).
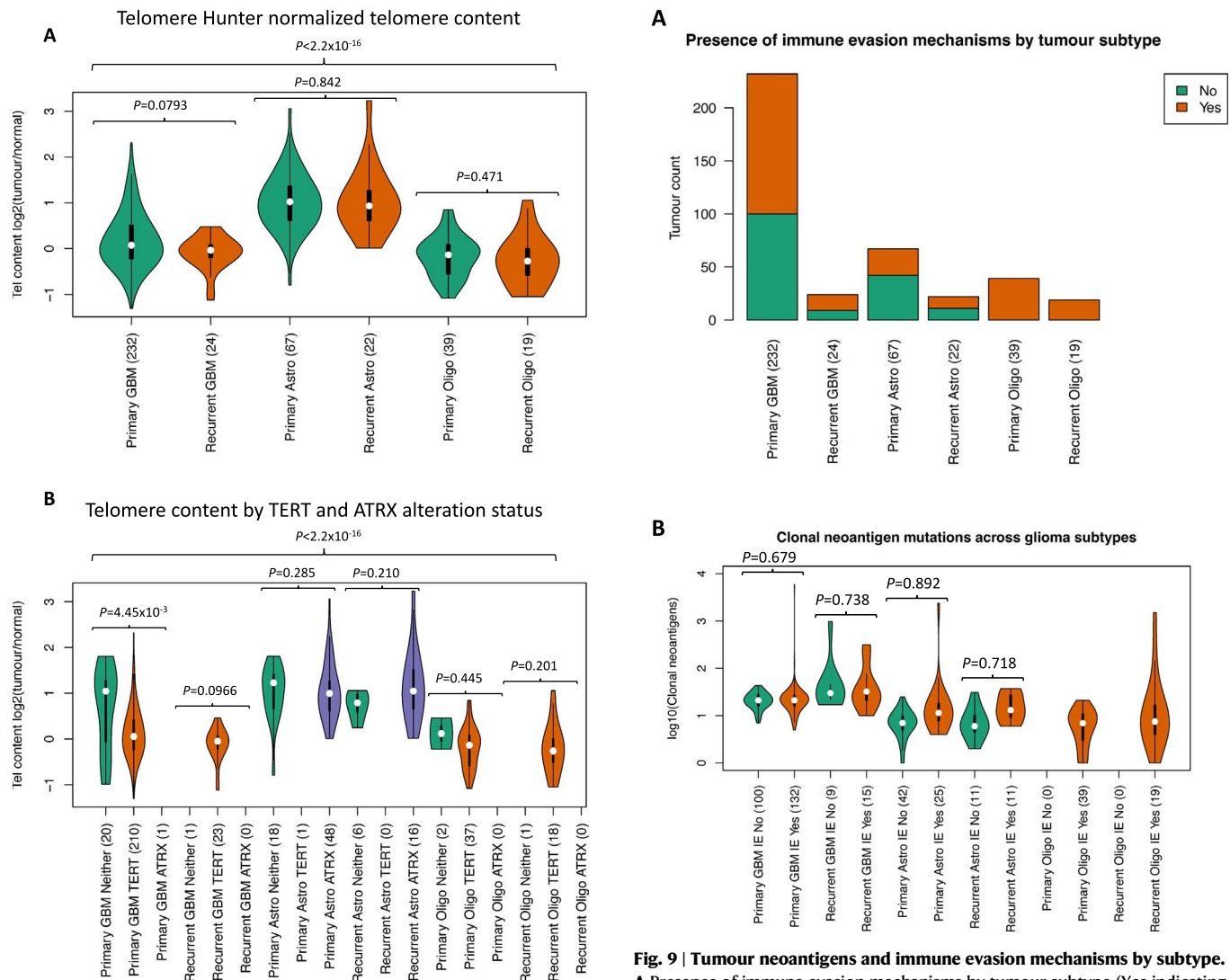
A

A Telomere Hunter normalized telomere content

B

Telomere content by TERT and ATRX alteration status

Fig. 8 | Telomere content by subtype and telomere maintenance mechanism.
A Telomere Hunter normalized telomere content by glioma subtype; **B** Telomere Hunter normalized telomere content by TERT and ATRX alteration status. Unadjusted two-sided *P*-values estimated from Kruskal-Wallis test between indicated tumour subsets with sample numbers indicated in brackets. In the box plots, the centre line represents the median, and the box bounds represents the inter-quartile range. Total *n* = 403.

A

Presence of immune evasion mechanisms by tumour subtype

B

Clonal neoantigen mutations across glioma subtypes

Fig. 9 | Tumour neoantigens and immune evasion mechanisms by subtype.
A Presence of immune evasion mechanisms by tumour subtype (Yes indicating tumours with at least one immune evasion mechanism, otherwise No); **B** Clonal neoantigen mutations across glioma subtypes and presence of immune evasion mechanism. Unadjusted two-sided *P*-values estimated from Kruskal-Wallis test between indicated tumour subsets with sample numbers indicated in brackets. In the box plots, the centre line represents the median, and the box bounds represents the inter-quartile range. Total *n* = 403.

Examining for evidence of immune evasion we considered loss of heterozygosity (LOH) or mutation of HLA class I genes and immune escape genes (IEGs; see Methods). 57% of GBM tumours showed evidence of at least one mechanism compared with 40% of astrocytomas and 100% of oligodendrogliomas (Fig. 9; Supplementary Data 26). While mutation or LOH of HLA and IEG mutations were infrequent events, the most common basis of immune evasion in all tumours was IEG LOH (Supplementary Fig. 27). There was no significant difference between the proportion of clonal neoantigens in carriers versus non carriers of immune evasion mechanisms in any tumour type, arguing against neoantigen production being under considerable selective constraint in glioma (Fig. 9; Supplementary Data 26).

Prognostic insights
We examined for associations between molecular and clinical features in primary and recurrent glioma subtypes (Supplementary Data 4, 16, 24, and 27–31). Notable findings include a significant association between *MGMT* methylation status and *CDKN2A* loss in primary GBM

(Supplementary Data 16 and 28) that is negative for loss of one copy of *CDKN2A* (P_{fisher} = 0.02; OR = 0.42) but positive for *CDKN2A* homozygous deletion (HOMDEL; P_{fisher} = 0.0031; OR = 2.3). EGFRvIII mutation appeared to be mutually exclusive with gain or mutation of *PDGFRA* (P_{fisher} = 0.028; OR = 0) and gain of chr17q (P_{fisher} = 0.01; OR = 0), but was positively associated with gain of *MDM4* (P_{fisher} = 0.0073; OR = 4.2) in primary GBM (Supplementary Data 16 and 28).

Notable associations with worse survival, as per previous reports, include *TP53* mutation[61] (P_{coxph} = 0.04; HR = 1.40, 95% CI 1.02–1.94), *CDKN2A* LOH (P_{coxph} = 0.04; HR = 1.49, 95% CI 1.02–2.19) and chr22q LOH (P_{coxph} = 0.04; HR = 1.39, 95% CI 1.02–1.90) in primary GBM[62], *PIK3CA* mutations (P_{coxph} = 0.005; HR = 9.04, 95% CI 1.92–42.51) and *TP53* 3'UTR germline variant rs7837822 (P_{coxph} = 0.0089; HR = 7.88, 95% CI 1.68–37.03) in primary astrocytoma, SBS40 signature presence (P_{coxph} = 0.02; HR = 8.57, 95% CI 1.36–54.03), immune evasion gene LOH (P_{coxph} = 0.03; HR = 5.67, 95% CI 1.23–26.14) and ID8 signature presence (P_{coxph} = 0.04; HR = 6.08, 95% CI 1.07–34.51) in recurrent astrocytoma and *BCOR* mutation (P_{coxph} = 0.03; HR = 14.60, 95% CI

1.21–175.59) in recurrent oligodendroglioma[63] (Fig. 4; Supplementary Data 15; Supplementary Fig. 3).

Actionable alterations

Systematic analyses of cancer genomes provide an opportunity of estimating the number of patients eligible for a targeted therapy and identify potentially novel therapeutic interventions. We annotated genomic alterations for clinical actionability using OncoKB Knowledge Base v3.17[64], which catalogues approved drugs having demonstrated efficacy based on clinical trials and published clinical evidence. OncoKB also provides compelling biological evidence supporting the cancer driver gene as being predictive of a response to a drug.

Based on OncoKB, the median number of targetable alterations in GBM tumours was four, while this was two in both astrocytoma and oligodendroglioma (OncoKB Level 1–4). Most mutations were Level 3B, which are predictive biomarkers that are considered standard-of-care in a different tumour type (Fig. 10; Supplementary Fig. 28; Supplementary Data 32). Example therapies for actionable alterations include ivosidenib for oncogenic *IDH1* mutations in astrocytomas and oligodendrogliomas[65,66], PI3K inhibitors in PTEN-deficient GBM tumours[67,68], and pembrolizumab in both TMB-H tumours[69] and MSI-H tumours[70–73] (Fig. 10).

We further identified 26 coding driver genes that are not currently catalogued by OncoKB, raising the prospect that these genes may be viable targets for novel therapeutic intervention (Supplementary Data 11). Based on DepMap[26] data, 14 of the 26 genes are predicted to be essential with positive selectivity, two being targetable with a pre-existing compound available (*CREBBP* and *PBRM1*), and three possessing a druggable structure or are druggable by ligand-based assessment (*PTPN11*, *FUBP1* and *SETD2*; Supplementary Data 11). Based on driver alteration frequencies in our study, up to 139 patients (34%) carried a targetable alteration for such a proposed novel therapy (Supplementary Data 11).

Discussion

Herein, we provide a detailed overview of the genomic landscape of adult diffuse glioma. Our study extends the number of such whole genomes reported by approximately four-fold, further defining the genomic basis of molecular subtypes and describing their evolutionary trajectories. We acknowledge that there are limitations to our study including the lack of functional data and the need for validation. Our reliance on short-read sequencing and lack of transcriptome information[74] to validate gene-fusions and functionally interrogate novel candidate cis-regulatory elements is a further limitation. However, as well as confirming established drivers we have been able to extend the genomic catalogue of glioma drivers within each subtype, informing them of their etiological basis and tumour vulnerability. Furthermore, by virtue of the large size of our study and the fact that within the UK NHS system patients are managed in a consistent fashion we have been able to explore the relationship between molecular features and patient outcomes in greater depth than previous reports.

Based on OncoKB several of the drivers we identify potentially represent attractive candidates a priori. For example, the potential targeting of SWI/SNF pathway genes. In surveying the non-coding landscape, as well as confirming the central role of *TERTp* mutations we found evidence of positive selection of mutations in several cis-regulatory elements, albeit at low frequencies, highlighting essential cell regulatory processes as potential targets for intervention. Future studies are important to provide experimental validation of our findings.

It has previously been purported[51] that ecDNA confers a worse survival. When appropriately restricted to GBM, we found no evidence to support such an assertion. It has also been suggested that ecDNA leading to EGFR amplification in GBM mediates resistance to EGFR inhibitors[52] and that ecDNA in general can mediate treatment-resistance mechanisms[53]. In our analysis there was no support for ecDNA conferring a selective advantage in tumours treated with the "standard" regimen (i.e. radiotherapy and temozolomide). Furthermore, no overrepresentation of ecDNA was shown in recurrent tumours ($P = 0.52$; OR = 0.72; Supplementary Data 16).

Our mutational signatures analysis suggests it is likely most driver mutations occur late and are acquired during progression to primary disease from clock-like mutational processes. Furthermore, while recurrent tumours were enriched for treatment-related signatures SBS11 and ID8, we found limited support for treatment-related driver mutations. While a strength of our study is the availability of primary and recurrent tumour samples, the vast majority are not matched. This limits the viability of conclusions relating to potential treatment-related driver mutations and mutational signatures, as in practically all cases the untreated primary sample is unavailable for comparison. Additional matched cohorts therefore would be required to follow-up on putative treatment-related mutations in *RB1* and *SMARCA4* in recurrent GBM and *SETD2* and *CREBBP* in recurrent oligodendroglioma identified in this study.

High levels of regulatory CD4+ tumour-infiltrating lymphocytes (TILs) and low levels of cytotoxic CD8+ TILs have been linked to a poorer outcome in glioma[75] and there is a growing consensus of the importance of the tumour microenvironment in mediating response to treatment and tumour recurrence[76]. While recurrent tumours showed a higher predicted neoantigen burden, predicted immune escape was a common feature irrespective of subtype. This may go some way to explain in part the variable response to immune checkpoint inhibition[77]. This should not detract from such agents having clinical benefit in patients with germline mutations in *MLH1* or *POLE* (Supplementary Data 23) whose tumours will be resistant to alkylating agent chemotherapies, such as temozolomide[78,79].

In conclusion, our study sheds light on the diversity of genomic processes driving the development of each subtype of glioma, oncogenesis and progression, and suggests targets for potential therapeutic intervention.

Methods

Ethics statement

Whole-genome sequencing (WGS) for glioma was performed through the 100,000 Genomes project approved by the East of England – Cambridge South Research Ethics Committee (REC reference[14]/EE/1112). All patients provided written informed consent and the study was conducted in accordance with the Declaration of Helsinki (see https://www.genomicsengland.co.uk/initiatives/100000-genomes-project/documentation for further information on patient consent and withdrawal). Sex-stratified analysis was not considered for this study, and there are no results that are only applicable to a single sex.

Participants and glioma samples

The research was conducted within the Research Environment (a secure data platform complying with the Department of Health Information Governance policy and meeting the data security standards of the Information Governance Toolkit of the Health and Social Care Information Centre) using LabKey (Main Programme Version 8 release, 28/11/2019). We excluded samples with either high sample cross-contamination or low tumour sample purity. Tumour sequence information was integrated into glioma tumour classification according to the CNS WHO v5 criteria[13]. Samples were prepared using an Illumina TruSeq DNA library preparation kit and sequenced on a HiSeq X, generating 150 base pair (bp) paired-end reads. Tumour and constitutional DNA samples were sequenced to average depths of 100x and 33x, respectively. Poor sequencing quality outliers were identified using principal component analysis and removed based on the following quality metrics: percentage of mapped reads; percentage of chimeric DNA fragments; average insert size; AT/CG dropout; and
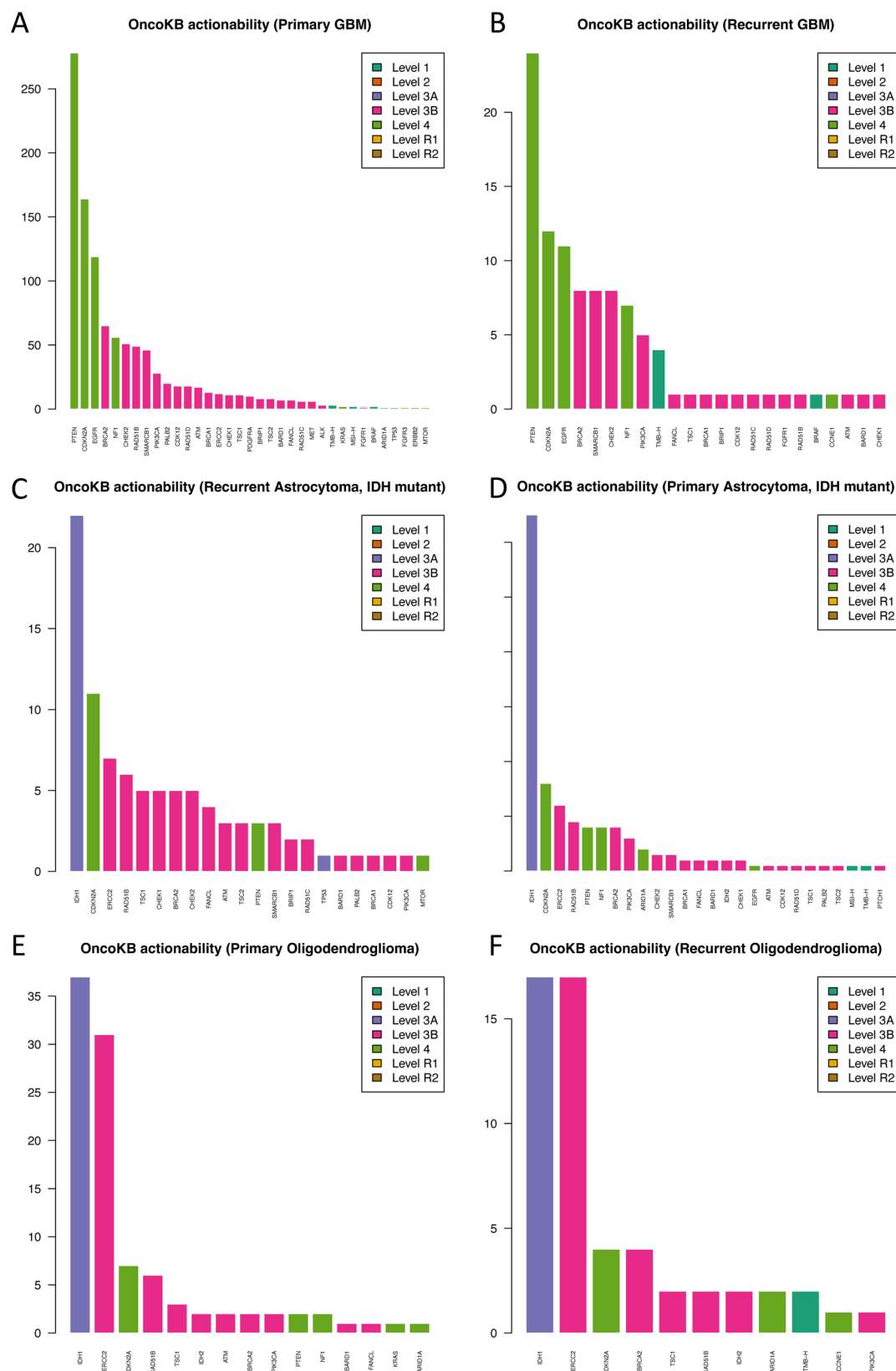
Fig. 10 | Per-gene actionable events catalogued by OncoKB. Level 1: FDA-recognised biomarker predictive of response to an FDA-approved drug in this indication; Level 2: Standard care biomarker predictive of a response to an FDA-approved drug in this indication; Level 3A: Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication; Level 3B: Standard care or investigational biomarker predictive of response to an FDA-approved or investigational drug in another indication; Level 4: Compelling biological evidence supports the biomarker as being predictive of response to a drug. **A** Primary GBM; **B** recurrent GBM; **C** primary astrocytoma; **D** recurrent astrocytoma; **E** primary oligodendroglioma; **F** recurrent oligodendroglioma.

unevenness of local coverage. Illumina's North Star pipeline (v2.6.53.23) was used for the primary whole-genome sequencing analysis. Sequence reads were aligned to the *Homo sapiens* GRCh38Decoy assembly using Isaac (v03.16.02.19). After the outlined quality control steps, tumour and germline sequencing data for 403 fresh-frozen adult glioma samples were obtained from the 100 kGP main programme version 8 release and taken into our analysis (Supplementary Data 33).

Variant calling and filtering

Single-nucleotide variant (SNV) and small insertion and deletion (indel) calling were performed using Strelka (v2.4.7). In addition to the default Strelka filters, we applied further filters as described in Supplementary Methods. The Isaac --clip-semialigned parameter invokes the soft clipping of read ends until five consecutive bases are matched with the reference genome[80]. This soft clipping therefore results in the loss of support for alternate alleles occurring within five bases of each read end, leading to artefactually low VAFs. To address allelic bias introduced by this clipping, we introduced FixVAF to soft clip all reads by 5 bp at each end, regardless of whether any of the bases are variant sites or whether the reads support reference or alternate alleles. Reads containing small insertions and deletions at variant positions were ignored.

Mutational signature analysis

Single base substitution (SBS), doublet base substitutions (DBS) and small insertions and deletions (ID) signatures were inferred from filtered Strelka input VCFs using SigProfilerMatrixGenerator v1.1.30[81] and SigProfilerExtractor v1.1.4[54]. Signatures were extracted de novo and decomposed to known COSMIC signatures (v3.2). The following parameters were used: "minimum_signatures = 1, maximum_signatures = 25, nmf_replicates = 500, nmf_init='random', matrix_normalization='gmm', min_nmf_iterations = 10000, max_nmf_iterations = 1000000, nmf_test_conv = 10000", otherwise default parameters were used.

Copy number alteration profiling

Clonal and sub-clonal somatic copy number alterations (CNAs) were detected using an iterative procedure incorporating Battenberg v2.2.8[20] (Supplementary Figs. 29–30). This procedure comprises four stages: (stage 1) initial CNA profiling; (stage 2) evaluation of CNA profile concordance with variant allele frequency (VAF) distributions; (stage 3) quality assessment; and for those samples that fail quality assessment, (stage 4) CNA re-profiling with alternative purity and ploidy estimates and repeat of stages 2 and 3. Further detailed description in Supplementary Methods.

Structural variants

We identified structural variants (SVs; also referred to as rearrangements) using a graph-based consensus approach including Delly[23], Lumpy[22] and Manta[21], and support from CNAs. Delly was run with post-filtering of somatic SVs using all normal samples, and Lumpy and Manta were run with default parameters. Rearrangements from the three SV callers were excluded if <2% of tumour reads at the rearrangement breakpoint site supported the rearrangement, or if any reads in the matched normal supported the rearrangement, or if either rearrangement breakpoint was in a centromeric or telomeric region, or on a non-standard reference contig (not chromosomes 1–22, X or Y). Remaining rearrangements were merged using a modified version of PCAWG Merge SV, which is a graph-based approach that identifies and merges rearrangements from multiple callers, allowing 400 bp slop for ambiguity in rearrangement breakpoint position[18]. Rearrangements were included in the final call set if they were identified by at least two SV callers, or by one SV caller but with a rearrangement breakpoint <3 kb from a CNA segment boundary. Please see detailed description in Supplementary Methods

on SV classification, hotspots of simple SVs, genomic regions enriched for chromothripsis and SV signatures.

Neoantigen and immune escape prediction

HLA alleles were detected from germline bam files using HLA-LA v1.0.1[82] (https://github.com/DiltheyLab/HLA-LA). Input somatic and phased germline-somatic VCFs were prepared for neoantigen detection as described (https://pvactools.readthedocs.io/en/latest/pvacseq/input_file_prep/proximal_vcf.html) in order to correct for potential germline effects on candidate peptides/neoantigens[83]. pVACseq v1.5.4[84] (https://pvactools.readthedocs.io/en/latest/pvacseq.html) was used to identify candidate neoantigens. Default parameters were used in filtering and identification of candidate neoantigens (i.e. binding <500 nM). Detection of allelic imbalance and loss of heterozygosity at HLA-A, HLA-B and HLA-C alleles was carried out using LOHHLA[85] (https://bitbucket.org/mcgranahanlab/lohhla/src/master/).

Detection of recurrent broad and focal copy number alterations

For every tumour passing copy number quality control measures, a copy number segmentation file was generated as input for GISTIC from Battenberg per-tumour segmentation output. From each copy number segment identified by Battenberg, the chromosomal coordinates, major (nMaj) and minor (nMin) copy number calls were obtained. In the case of subclonal copy number segments, nMaj and nMin were taken from the subclone with the largest tumour fraction. GISTIC v2.0.2.3[38] was run to identify recurrent arm-level copy number events, as well as focal amplifications and deletions (https://github.com/broadinstitute/gistic2). The following parameters were used: "-conf 0.99 -broad 1 -qvt 0.25 -genegistic 1 -gcm extreme -brlen 0.5 -rx 0 -twoside 1 -scent median -armpeel 1 -arb 1 -refgene hg38.UCSC.add_-miR.160920.refgene.mat". Candidate target genes at focal amplifications and deletions were annotated according to the criteria mentioned in Supplementary Methods.

Telomere analysis methods

Telomere content was estimated from tumour and germline bam files by the methods Telomere Hunter v1.1.0[86] (https://pypi.org/project/telomerehunter/) and Telomerecat v3.3.0[87] (https://github.com/cancerit/telomerecat) using default parameters. Telomere content was normalized by (1) $\log 2(\frac{\text{Tumour}}{\text{Germline}})$. Telomere content estimates were compared between methods as well as across age groups, tumour purity and WGD status (Supplementary Figs. 31–35).

Extrachromosomal DNA detection

Potential extrachromosomal DNA (ecDNA) molecules were detected from tumour bam files using AmpliconArchitect v1.2[49] (https://github.com/virajbdeshpande/AmpliconArchitect). Briefly, per-tumour "seed" regions were prepared from Battenberg copy-number segmentation output if a segment was >100 kb and the total copy number was >5. AmpliconArchitect was then run using these "seed" regions to extract overlapping sequence reads from the tumour bamfile and construct candidate amplicons.

Candidate amplicons were classified using AmpliconClassifier v0.4.6 (https://github.com/jluebeck/AmpliconClassifier) into the following categories: (1) Cyclic (truly circularised ecDNA), (2) Complex non-cyclic, (3) Linear amplification, (4) No amp/invalid. Amplicons were highlighted if containing a known highly amplified oncogene (*MDM2, MYC, EGFR, CDK4, ERBB2, SOX2, TERT, CCND1, E2F3, CCNE1, CDK6, MDM4, NEDD9, MCL1, AKT3, BCL2L1, ZNF217, KRAS, PDGFRA, AKT1, MYCL, NKX2-1, IGF1R, PAX8*; as per Kim et al., 2020).

Annotation of consequence of structural variants

The final set of high-quality structural variant (SV) calls were annotated by SNPeff v5.1[88] (http://pcingola.github.io/SnpEff/) according to GRCh38 ensembl v101 reference to identify their likely effect on gene

transcript and/or protein function. Annotations were restricted to canonical protein-coding transcripts in ensembl v101. These annotated SV calls were further prioritised using "simple sv annotation" (https://github.com/AstraZeneca-NGS/simple_sv_annotation) to highlight exon-level deletion events in candidate driver genes.

Driver analysis methods
Somatic mutations were annotated to GRCh38 ensembl v101 using the variant effect predictor[89] (VEP). The following parameters were used: "vep -i <input_vcf> --assembly GRCh38 −no_stats −cache −offline −symbol −protein -o <output> --vcf −canonical −dir <ref_dir> --hgvs −hgvsg −fasta <GRCh38_fasta> --plugin CADD, <CADD_score_file> --plugin UTRannotator,<GRCh38_uORF_reference>". The <CADD_score_file> was obtained using CADD v1.6[90–92] (https://cadd.gs.washington.edu/) with scores obtained for all SNV and indel mutations using the CADD software (https://github.com/kircherlab/CADD-scripts/), before being utilised by the VEP CADD plugin. The plugin "UTRannotator"[93] (https://github.com/ImperialCardioGenetics/UTRannotator) was used to annotate the potential impact of five prime untranslated region (5′ UTR) mutations.

Protein-coding driver genes were identified using the IntOGen pipeline[24] (https://bitbucket.org/intogen/intogen-plus/src/master/; downloaded February 2021; https://intogen.readthedocs.io/en/latest/). Regions from candidate non-coding elements overlapping coding sequence (CDS) or exon regions from canonical protein-coding transcripts were removed using bedops v2.4.39[94] (https://github.com/bedops/bedops). SNV mutations that otherwise passed filtering criteria (Supplementary Methods) were further scrutinised by whether they exhibited excessive strand bias (Strelka INFO field "SNVSB="> 10). Nonsynonymous mutations in the 682 gene transcripts considered by OncoKB v3.3 were annotated using the OncoKB API (https://www.oncokb.org/)[64]. OncoKB v3.17 was used to annotate clinically actionable genomic alterations. Further details are provided in Supplementary Methods.

Timing driver mutations
The relative evolutionary timing of candidate driver mutations was obtained using MutationTimeR[17] (https://github.com/gerstung-lab/MutationTimeR). Copy number input for MutationTimeR was prepared from Battenberg segmentation files, with the clonal frequency of each segment taken as the tumour purity. In the case of subclonal calls, the clonal frequency was calculated by multiplying the tumour purity by the clonal fraction. The clusters input for MutationTimeR was prepared from DPClust cluster estimates. The VAF proportion was calculated by multiplying the estimated cluster CCF by the tumour purity. Superclonal clusters (CCF > 1.1) were removed. VCF input for MutationTimeR was obtained from the filtered small somatic SNV/indel variant VCFs (Supplementary Methods). For SNVs, alt and ref depths were obtained using FixVAF (https://github.com/danchubb/FixVAF). For indels, ref and alt depths were obtained from Tier2 Strelka TAR and TIR fields, respectively.

Mutational signature attribution
SeqInfo VCFs produced as part of SigProfilerMatrixGenerator (https://github.com/AlexandrovLab/SigProfilerMatrixGenerator) were used to map somatic mutations from input VCFs to their SBS96, DBS78 or ID83 contexts and then to the final SigProfilerExtractor COSMIC v3.2 decomposed signature probabilities. Dinucleotide mutations called originally by Strelka as adjacent SNVs were re-annotated by VEP. When deriving timing estimates from MutationTimeR the timing from one of the two adjacent SNVs was taken for a given dinucleotide mutation. Per-tumour VCFs containing DBS mutations, either directly called originally by Strelka, or originally called by Strelka as two adjacent SNVs and reconstructed as DBS mutations, were created and mutation consequences re-calculated using VEP as above.

Druggability and essentiality analysis of driver genes
Proteins of identified driver genes were annotated with pharmacological and druggability data using canSAR's Cancer Protein Annotation Tool (CPAT)[95]. Essential and selective genes were ascertained from the ShinyDepMap[26] analysis, using efficacy and selectivity score thresholds of <−0.56 and >0, respectively.

Clinical data collection and outcome measures
Participants from 20 neurosurgical units in the United Kingdom were included. The following demographic and operative parameters were captured: age, sex, date of surgery, tumour type, laterality tumour location, surgery type, post-operative complications (within 30 days), requirement for CSF diversion, WHO diagnosis/grade, histopathological report, molecular features (ATRX/IDH/1p19q/TERT/MGMT status (%)/p53), treatment (chemotherapy and radiotherapy (number of cycles, type), immunotherapy, trial participation, repeat or previous surgery), radiological extent of resection, treatment response (progression vs pseudoprogression vs stable disease), survival (last seen alive vs date of death), Eastern Cooperative Oncology Group performance status[96], any other comments.

Statistical analysis
Descriptive statistics were used to characterise the patient population. Statistical analysis was performed using GraphPad Prism v9 and The R Foundation for Statistical Computing Platform v4 statistical package. Chi-square test was used to assess the statistical significance of observed differences between different tumour cohorts. Kaplan-Meier plots were used to assess median overall survival and progression free survival for tumour subgroups. Cox proportional hazards regression was used to predict survival based on EGFR extrachromosomal DNA amplification (adjusted for age, sex and MGMT methylation status). False discovery rate (FDR) q-values were calculated using the Benjamini-Hochberg method unless otherwise stated.

Correlating genomic and clinical features
Key genomic and clinical features were extracted for pairwise statistical association testing. Each feature was categorized as either "Binary", "Categorical", "Survival" (in the case of overall survival and progression-free survival) or "Continuous" (Supplementary Data 31).

Pairwise comparisons between binary and categorical features were carried out using Fisher's exact test (python 3.9.19; FisherExact v1.4.2). Pairwise comparisons between binary/categorical and continuous features were carried out using the Kruskal-Wallis test (python 3.9.19; SciPy v1.13.1). Pairwise comparisons between continuous features were carried out using Spearman's rank correlation test (python 3.9.19; SciPy v1.13.1). Comparisons between genomic and clinical features with OS and PFS were carried out using cox's proportional hazards test (python 3.9.19; lifelines v0.29.0) with a step-size of 0.1, including the following covariates: patient age at diagnosis, patient sex.

Germline analysis
Germline-related analyses made use of the Aggregated Variant Calls (AggV2) resource provided by Genomics England. This briefly comprised an aggregated analysis of 78,195 germline genomes. Further details are available at Genomics England (https://re-docs.genomicsengland.co.uk/aggv2/). Of the 400 participants with diffuse glioma used in our landscape analysis, germline variant calls were available for 393.

Published code
Details of all reused published code and software are provided in Supplementary Data 35.

Reporting summary

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

Data availability

The data supporting the findings of this study are available within the Genomics England Research Environment, a secure workspace. All interested readers will be able to access the data in the same manner that the authors accessed the data. To access genomic and clinical data within the Research Environment, researchers must first apply to become a member of either the Genomics England Research Network (academics/healthcare professionals) or the Discovery Forum (industry) via the Genomics England website: https://www.genomicsengland.co.uk/research. The process for joining the Research Network consists of the following steps: 1. The researcher's institution will need to sign a Research Network Access Agreement. 2. Once confirmed the institution is registered with Genomics England, the researcher can apply for Research Network membership using the online form. 3. The institution will be asked to validate the researcher's affiliation with that institution and the application will then be reviewed by Genomics England within 10 working days. 4. The applicant completes online information governance training and will then be granted access to the Research Environment within 2 h of passing the online training. The processed clinical and genomic data used in this study as well as figure source data are available in the Research Environment within the folder /re_gecip/shared_allGeCIPs/glioma_landscape/results/. Germline variants can be explored in an Interactive Variant Analysis Browser, see description at: https://re-docs.genomicsengland.co.uk/iva_variant/. Further details of the types of data available in the Genomics England Research Environment can be found at: https://re-docs.genomicsengland.co.uk/data_overview/. There is no proposed end date for data access within the research environment. Details of all other datasets used in the study can be found in Supplementary Data 35.

Code availability

After completion of the instructions given by the Data availability statement, code to allow for reproducibility of results and figures are available in the research environment within the folder /re_gecip/shared_allGeCIPs/glioma_landscape/code/. Sources of each software package and externally downloaded data are shown in Supplementary Data 35.

References

1. Crocetti, E. et al. Epidemiology of glial and non-glial brain tumours in Europe. *Eur. J. Cancer* **48**, 1532–1542 (2012).
2. Bondy, M. L. et al. Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer* **113**, 1953–1968 (2008).
3. Koshy, M. et al. Improved survival time trends for glioblastoma using the SEER 17 population-based registries. *J. Neurooncol.* **107**, 207–212 (2012).
4. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
5. Eckel-Passow, J. E. et al. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N. Engl. J. Med.* **372**, 2499–2508 (2015).
6. Frattini, V. et al. The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* **45**, 1141–1149 (2013).
7. Kim, H. et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res.* **25**, 316–327 (2015).
8. Suzuki, H. et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).
9. Brat, D. J. et al. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
10. Noushmehr, H. et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17**, 510–522 (2010).
11. Sturm, D. et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
12. Verhaak, R. G. 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. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* **23**, 1231–1251 (2021).
14. Consortium, I.T.P.-C.A.O.W.G. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
15. Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111 (2020).
16. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
17. Gerstung, M. et al. The evolutionary history of 2658 cancers. *Nature* **578**, 122–128 (2020).
18. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121 (2020).
19. Cornish, A. J. et al. Reference bias in the Illumina Isaac aligner. *Bioinformatics* **36**, 4671–4672 (2020).
20. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
21. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
22. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
23. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
24. Martinez-Jimenez, F. et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
25. Wu, F. et al. Immunological profiles of human oligodendrogliomas define two distinct molecular subtypes. *EBioMedicine* **87**, 104410 (2023).
26. Shimada, K., Bachman, J. A., Muhlich, J. L. & Mitchison, T. J. shinyDepMap, a tool to identify targetable cancer genes and their functional connections from Cancer Dependency Map data. *Elife* **10**, e57116 (2021).
27. Rozenblatt-Rosen, O. et al. The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc. Natl Acad. Sci. USA* **106**, 755–760 (2009).
28. Chan, H. M. & La Thangue, N. B. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* **114**, 2363–2373 (2001).
29. Luo, R. et al. G protein-coupled receptor 56 and collagen III, a receptor-ligand pair, regulates cortical development and lamination. *Proc. Natl Acad. Sci. USA* **108**, 12925–12930 (2011).
30. Hale, C. F. et al. Essential role for vav Guanine nucleotide exchange factors in brain-derived neurotrophic factor-induced dendritic spine growth and synapse plasticity. *J. Neurosci.* **31**, 12426–12436 (2011).
31. Yao, X. et al. PBRM1-deficient PBAF complexes target aberrant genomic loci to activate the NF-κB pathway in clear cell renal cell carcinoma. *Nat. Cell Biol.* **25**, 765–777 (2023).

32. Michurina, A. et al. Postnatal expression of the lysine methyl-transferase SETD1B is essential for learning and the regulation of neuron-enriched genes. *EMBO J.* **41**, e106459 (2022).

33. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* **17**, 128 (2016).

34. Fishilevich, S. et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* https://doi.org/10.1093/database/bax028 (2017).

35. Hasan, M. M. et al. Disrupted neurogenesis, gliogenesis, and ependymogenesis in the Ccdc85c knockout rat for hydrocephalus model. *Cells Dev.* **175**, 203858 (2023).

36. Li, H. et al. Hsa_circ_0110757 upregulates ITGA1 to facilitate temo-zolomide resistance in glioma by suppressing hsa-miR-1298-5p. *Cell Death Dis.* **12**, 252 (2021).

37. Consortium, G. T. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

38. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).

39. Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

40. Melin, B. S. et al. Genome-wide association study of glioma sub-types identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* **49**, 789–794 (2017).

41. Nilsson, B., Johansson, M., Heyden, A., Nelander, S. & Fioretos, T. An improved method for detecting and delineating genomic regions with altered gene expression in cancer. *Genome Biol.* **9**, R13 (2008).

42. Glodzik, D. et al. A somatic-mutational process recurrently dupli-cates germline susceptibility loci and tissue-specific super-enhan-cers in breast cancers. *Nat. Genet* **49**, 341–348 (2017).

43. Le Tallec, B. et al. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **4**, 420–428 (2013).

44. Bignell, G. R. et al. Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).

45. Singh, D. et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**, 1231–1235 (2012).

46. Zheng, S. et al. A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. *Genes Dev.* **27**, 1462–1472 (2013).

47. Franklin, D. S. et al. CDK inhibitors p18(INK4c) and p27(Kip1) mediate two separate pathways to collaboratively suppress pituitary tumorigenesis. *Genes Dev.* **12**, 2899–2911 (1998).

48. Furgason, J. M. et al. Whole genome sequence analysis links chro-mothripsis to EGFR, MDM2, MDM4, and CDK4 amplification in glioblastoma. *Oncoscience* **2**, 618–628 (2015).

49. Deshpande, V. et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat. Commun.* **10**, 392 (2019).

50. Turner, K. M. et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).

51. Kim, H. et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat. Genet* **52**, 891–897 (2020).

52. Nathanson, D. A. et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).

53. Lange, J. T. et al. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nat. Genet* **54**, 1527–1533 (2022).

54. Islam, S. M. A. et al. Uncovering novel mutational signatures by *de novo* extraction with SigProfilerExtractor. *Cell Genom* **2**, None (2022).

55. Erson-Omay, E. Z. et al. Somatic POLE mutations cause an ultra-mutated giant cell high-grade glioma subtype with better prog-nosis. *Neuro Oncol.* **17**, 1356–1364 (2015).

56. Jin, S. G., Meng, Y., Johnson, J., Szabó, P. E. & Pfeifer, G. P. Con-cordance of hydrogen peroxide-induced 8-oxo-guanine patterns with two cancer mutation signatures of upper GI tract tumors. *Sci. Adv.* **8**, eabn3815 (2022).

57. Drost, J. et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **358**, 234–238 (2017).

58. Reijns, M. A. M. et al. Signatures of TOP1 transcription-associated mutagenesis in cancer and germline. *Nature* **602**, 623–631 (2022).

59. Vredenburgh, J. J. et al. Bevacizumab plus irinotecan in recurrent glioblastoma multiforme. *J. Clin. Oncol.* **25**, 4722–4729 (2007).

60. Li, B. et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood* **135**, 41–55 (2020).

61. Noor, H., Briggs, N. E., McDonald, K. L., Holst, J. & Vittorio, O. TP53 mutation is a prognostic factor in lower grade glioma and may influence chemotherapy efficacy. *Cancers* **13** (2021).

62. Lu, V. M. et al. The prognostic significance of CDKN2A homozygous deletion in IDH-mutant lower-grade glioma and glioblastoma: a systematic review of the contemporary literature. *J. Neurooncol.* **148**, 221–229 (2020).

63. Halani, S. H. et al. Multi-faceted computational assessment of risk and progression in oligodendroglioma implicates NOTCH and PI3K pathways. *NPJ Precis. Oncol.* **2**, 24 (2018).

64. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* **2017**, PO.17.00011 (2017).

65. Mellinghoff, I. K. et al. Ivosidenib in isocitrate dehydrogenase 1-mutated advanced glioma. *J. Clin. Oncol.* **38**, 3398–3406 (2020).

66. Popovici-Muller, J. et al. Discovery of AG-120 (Ivosidenib): a first-in-class mutant IDH1 inhibitor for the treatment of IDH1 mutant can-cers. *ACS Med. Chem. Lett.* **9**, 300–305 (2018).

67. Hansen, A. R. et al. A first in human phase I study of AZD8186, a potent and selective inhibitor of PI3K in patients with advanced solid tumours as monotherapy and in combination with the dual mTORC1/2 inhibitor vistusertib (AZD2014) or abiraterone acetate. *J. Clin. Oncol.* **35**, 2570–2570 (2017).

68. Mateo, J. et al. A first-time-in-human study of GSK2636771, a phosphoinositide 3 kinase beta-selective inhibitor, in patients with advanced solid tumors. *Clin. Cancer Res.* **23**, 5981–5992 (2017).

69. Marabelle, A. et al. 1192O - Association of tumour mutational burden with outcomes in patients with select advanced solid tumours treated with pembrolizumab in KEYNOTE-158. *Ann. Oncol.* **30**, v477–v478 (2019).

70. Abida, W. et al. Analysis of the prevalence of microsatellite instability in prostate cancer and response to immune checkpoint blockade. *JAMA Oncol.* **5**, 471–478 (2019).

71. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).

72. Le, D. T. et al. Phase II open-label study of pembrolizumab in treatment-refractory, microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: KEYNOTE-164. *J. Clin. Oncol.* **38**, 11–19 (2020).

73. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair defi-ciency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).

74. Mathur, R. et al. Glioblastoma evolution and heterogeneity from a 3D whole-tumor perspective. *Cell* **187**, 446–463.e16 (2024).

75. Han, S. et al. Tumour-infiltrating CD4(+) and CD8(+) lymphocytes as predictors of clinical outcome in glioma. *Br. J. Cancer* **110**, 2560–2568 (2014).

76. Hoogstrate, Y. et al. Transcriptome analysis reveals tumor micro-environment changes in glioblastoma. *Cancer Cell* **41**, 678–692.e7 (2023).

77. Touat, M. et al. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 (2020).

78. Germano, G., Amirouchene-Angelozzi, N., Rospo, G. & Bardelli, A. The clinical impact of the genomic landscape of mismatch repair-deficient cancers. *Cancer Discov.* **8**, 1518–1528 (2018).

79. Zhao, H. et al. Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks. *Elife* **3**, e02725 (2014).

80. Raczy, C. et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041–2043 (2013).

81. Bergstrom, E. N. et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).

82. Dilthey, A. T. et al. HLA*LA-HLA typing from linearly projected graph alignments. *Bioinformatics* **35**, 4394–4396 (2019).

83. Hundal, J. et al. Accounting for proximal variants improves neoantigen prediction. *Nat. Genet.* **51**, 175–179 (2019).

84. Hundal, J. et al. pVACtools: a computational toolkit to identify and visualize cancer neoantigens. *Cancer Immunol. Res.* **8**, 409–420 (2020).

85. McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271.e11 (2017).

86. Feuerbach, L. et al. TelomereHunter—in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinforma.* **20**, 272 (2019).

87. Farmery, J. H. R., Smith, M. L., Diseases, N. B.-R. & Lynch, A. G. Telomerecat: a ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Sci. Rep.* **8**, 1300 (2018).

88. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

89. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).

90. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet* **46**, 310–315 (2014).

91. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).

92. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **13**, 31 (2021).

93. Zhang, X., Wakeling, M., Ware, J. & Whiffin, N. Annotating high-impact 5'untranslated region variants with the UTRannotator. *Bioinformatics* **37**, 1171–1173 (2021).

94. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).

95. di Micco, P. et al. canSAR: update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res.* **51**, D1212–D1219 (2023).

96. Oken, M. M. et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am. J. Clin. Oncol.* **5**, 649–655 (1982).

Author contributions
J.J., D.C. and B.K. processed clinical data. A.J.C. performed sequencing data quality control. D.C. performed quality control of simple somatic mutations. A.J.C. and A.F. called copy number alterations. A.J.C. called structural variants and identified recurrent events. B.K. identified recurrent copy number alterations. B.K. identified driver mutations. B.K., A.J.C., D.C. and A.J.G. extracted mutational signatures. B.K. carried out HLA estimation, neoantigen prediction and immune escape prediction. B.K. and S.L. carried out LOHHLA analysis. B.K. estimated telomere content. B.K. inferred driver mutation clonality. B.K. correlated genomic and clinical features. J.J. and B.K. carried out survival analysis. A.Sud and B.K. performed clinical actionability analyses. R.S.H. and K.A. supervised the study. B.K., J.J., A.J.C., D.C. and R.S.H. wrote the manuscript. All authors (B.K., J.J., A.J.C., D.C., R.L., A.F., A.J.G., A.Sud, G.C., A.Sottoriva, D.C.W., T.B., S.A-S., S.E.D.L., E.A., G.A., D.B., A.B., Y.A.C., P.D'U., R.C., A.D., E.E., F.G., P.G., C.S.H., M.D.J., S.T.K., S.Smith, S.Manivannan, A.J.M., S.Matloob, S.Mukherjee, K.O'N., P.P., J.P., S.P., O.R., B.S., F.S., S.Sinha, L.T., I.U., P.W., A.S.Y., H.B., P.A., R.H. and K.A.) read and approved the final version of this manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59156-9.

Correspondence and requests for materials should be addressed to Ben Kinnersley, Josephine Jung, Richard Houlston or Keyoumars Ashkan.

Peer review information *Nature Communications* thanks Anton Henssen, Alvin Wei Tian Ng, Derek Oldridge, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London SM2 5NG, UK. [2]UCL Cancer Institute, 72 Huntley St, WC1E 6DD London, UK. [3]Institute of Psychiatry, Psychology and Neurosciences, Kings College London, Strand, WC2R 2LS London, UK. [4]Department of Neurosurgery, King's College Hospital NHS Foundation Trust, Denmark Hill, SE5 9RS London, UK. [5]Department of Clinical Neuropathology, King's College Hospital NHS Foundation Trust, Denmark Hill, SE5 9RS London, UK. [6]Cancer Genomics, Big Data Institute, Nuffield Department of Medicine, Old Road Campus, OX3 7LF Oxford, UK. [7]Department of Biology, University of Konstanz, Konstanz 78464, Germany. [8]Evolutionary Genomics and Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London SM2 5NG, UK. [9]Manchester Cancer Research Centre, University of Manchester, 555 Wilmslow Road, Manchester M20 4GJ, UK. [10]School of Biomedical Engineering & Imaging Sciences, King's College London, St Thomas's Hospital, London, UK. [11]Department of Neuroradiology, King's College Hospital NHS Foundation Trust, Denmark Hill, SE5 9RS London, UK. [12]Department of Neurosurgery, Royal Stoke University Hospital, Newcastle Road, ST4 6QG Stoke-on-Trent, UK. [13]Imperial College Healthcare NHS Trust, Charing Cross Hospital, 3S corridor, Fulham Palace Road, London W6 8RF, UK. [14]Department of Neurosurgery, Royal National Orthopaedic Hospital, Brockley Hill, Stanmore HA7 4LP, UK. [15]Department of Neurosurgery, John Radcliffe Hospital, Headley Way, Headington, OX3 9DU Oxford, UK. [16]Department of Neurosurgery, Queen Elizabeth Hospital Birmingham, Mindelsohn Way, B15 2GW Birmingham, UK. [17]Department of Neurosurgery, Manchester Royal Infirmary, Oxford Rd, M13 9WL Manchester, UK. [18]Department of Neurosurgery, Leeds General Infirmary, Great George St, LS1 3EX Leeds, UK. [19]Department of Neurosurgery, Queen's Medical Centre NHS Trust, Derby Road, Lenton, NG7 2UH Nottingham, UK. [20]South West Neurosurgery Unit, University Hospitals Plymouth NHS Trust, Derriford Road, Crownhill, PL6 8DH Plymouth, UK. [21]Department of Neurosurgery, St. George's University Hospitals NHS Foundation Trust, Blackshaw Rd, SW17 0QT London, UK. [22]Department of Neurosurgery, Southampton General Hospital, Tremona Road, SO16 6YD Southampton, UK. [23]Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, Queen Square, WC1N 3BG London, UK. [24]Department of Neurosurgery, The Walton Centre NHS Foundation Trust, Lower Lane, Fazakerley, L9 7LJ Liverpool, UK. [25]Department of Neurosurgery, Queen's Hospital Romford, Rom Valley Way, RM7 0AG Romford, UK. [26]Department of Neurosurgery, Addenbrookes Hospital, Hills Rd, CB2 0QQ Cambridge, UK. [27]Department of Neurosurgery, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, Glossop Rd, Broomhall, S10 2JF Sheffield, UK. [28]Brainstrust, 4 Yvery Court, Castle Road, PO31 7QG Cowes, Isle of Wight, UK. [29]Genomics England, One Canada Square, London E14 5AB, UK. [30]These authors contributed equally: Ben Kinnersley, Josephine Jung, Alex J. Cornish, Daniel Chubb. [31]These authors jointly supervised this work: Richard Houlston, Keyoumars Ashkan. ✉e-mail: b.kinnersley@ucl.ac.uk; josephine.1.jung@kcl.ac.uk; richard.houlston@icr.ac.uk; k.ashkan@nhs.net