nature genetics

Spatiotemporal genomic architecture informs precision oncology in glioblastoma

Jin-Ku Lee[1,2,13], Jiguang Wang[3–5,13], Jason K Sa[1,2,6,13], Erik Ladewig[3,4,13], Hae-Ock Lee[7], In-Hee Lee[1,2], Hyun Ju Kang[1,2], Daniel S Rosenbloom[3,4], Pablo G Camara[3,4], Zhaoqi Liu[3,4], Patrick van Nieuwenhuizen[3,4], Sang Won Jung[1,2,6], Seung Won Choi[1,2,6], Junhyung Kim[1,2], Andrew Chen[3,4], Kyu-Tae Kim[7], Sang Shin[1,2,6], Yun Jee Seo[1,2], Jin-Mi Oh[1,2], Yong Jae Shin[1,2,7], Chul-Kee Park[8], Doo-Sik Kong[2], Ho Jun Seol[2], Andrew Blumberg[9], Jung-Il Lee[2], Antonio Iavarone[10–12], Woong-Yang Park[6,7], Raul Rabadan[3,4] & Do-Hyun Nam[1,2,6]

Precision medicine in cancer proposes that genomic characterization of tumors can inform personalized targeted therapies[1–5]. However, this proposition is complicated by spatial and temporal heterogeneity[6–14]. Here we study genomic and expression profiles across 127 multisector or longitudinal specimens from 52 individuals with glioblastoma (GBM). Using bulk and single-cell data, we find that samples from the same tumor mass share genomic and expression signatures, whereas geographically separated, multifocal tumors and/or long-term recurrent tumors are seeded from different clones. Chemical screening of patient-derived glioma cells (PDCs) shows that therapeutic response is associated with genetic similarity, and multifocal tumors that are enriched with *PIK3CA* mutations have a heterogeneous drug-response pattern. We show that targeting truncal events is more efficacious than targeting private events in reducing the tumor burden. In summary, this work demonstrates that evolutionary inference from integrated genomic analysis in multisector biopsies can inform targeted therapeutic interventions for patients with GBM.

Most clinical trials for targeted therapy in GBM have shown limited clinical success[15]. Although recent genome-wide studies evaluating regional heterogeneity[9,11] and longitudinal GBM pairs[8,12,14,16] have suggested potential evolutionary models of the tumors, there is little understanding regarding which strategies can effectively use genomic data to inform targeted therapies. To identify such strategies, we analyzed somatic variants in 127 multiregion or longitudinal tumor specimens from 52 individuals with glioma: 42 individuals from the Samsung Medical Center (SMC), Seoul, and 10 individuals from The Cancer Genome Atlas (TCGA) GBM cohort[8,14,16] (**Supplementary Table 1**). Additionally, we analyzed the transcriptomes of 83 tumor specimens from 41 individuals (bulk) and 305 single cells from 7 samples for 3 individuals. Tumors were classified into four distinct groups according to the spatial and temporal features of tissue acquisition: tumors obtained from the same location at the same time (locally adjacent), tumors obtained from different locations at the same time (multifocal/multicentric; referred to as multiple)[17], and tumors obtained from local and distant recurrences at different times (longitudinal local and distant, respectively) (**Fig. 1a**).

We inferred clonal and subclonal alterations from cancer cell fractions in multiple sectors (Online Methods and **Supplementary Table 2**). The average mutation rate was 2.2 mutations/Mb for non-hypermutated samples, which is consistent with previous studies[2,14]. *IDH1* mutations mapping to Arg132 were clonal across all of the regions in IDH-mutant tumors[2,14] (6/6). *PIK3CA* mutations were always clonal and shared by all sectors (5/5), which is consistent with the results of our previous longitudinal analysis from tumor evolution directed graphs (TEDGs), in which we found that *PIK3CA* mutations are early events[14] (**Fig. 1b** and **Supplementary Fig. 1a**,**b**). Furthermore, *FGFR3-TACC3* fusions were highly expressed in all regions from two individuals[18]. These somatic variants, which are shared by all tumor regions, represent promising therapeutic targets[14,19], as they reflect truncal alterations in the evolutionary tree that are suspected to be present among all tumor cells. In contrast, *PTEN* alterations, including copy number deletions and mutations, were shared by 10 of 20 (50%) and 5 of 7 (71.4%) tumor sectors, respectively. Likewise, *EGFR* amplifications were observed as private events, exclusive to only one or two regions of the multisector

[1]Institute for Refractory Cancer Research, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [2]Department of Neurosurgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [3]Department of Systems Biology, Columbia University, New York, New York, USA. [4]Department of Biomedical Informatics, Columbia University, New York, New York, USA. [5]Divisions of Life Science and Biomedical Engineering, Hong Kong University of Science and Technology, Hong Kong. [6]Department of Health Science and Technology, Samsung Advanced Institute for Health Science and Technology, Sungkyunkwan University, Seoul, Republic of Korea. [7]Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [8]Department of Neurosurgery, College of Medicine, Seoul National University and Seoul National University Hospital, Seoul, Republic of Korea. [9]Department of Mathematics, University of Texas, Austin, Texas, USA. [10]Institute for Cancer Genetics, Columbia University, New York, New York, USA. [11]Department of Neurology, Columbia University, New York, New York, USA. [12]Department of Pathology, Columbia University, New York, New York, USA. [13]These authors contributed equally to this work. Correspondence should be addressed to R.R. (rr2579@cumc.columbia.edu), W.-Y.P. (woongyang@skku.edu) or D.-H.N. (nsnam@skku.edu).

Received 1 August 2016; accepted 10 February 2017; published online 6 March 2017; doi:10.1038/ng.3806

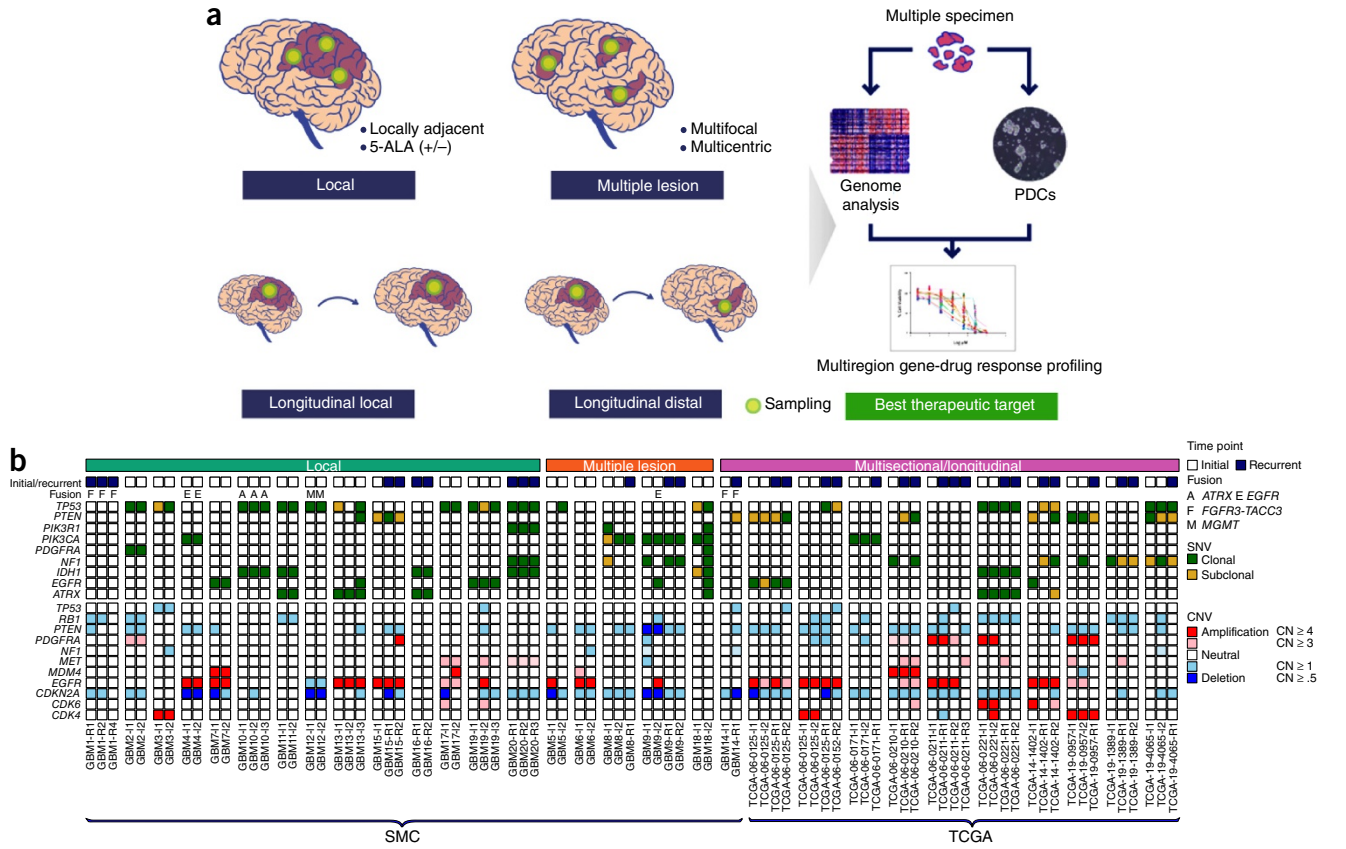

Figure 1 Mutational landscape of multiregion malignant glioma samples. (**a**) A schematic representation of glioma genomic heterogeneity and differential drug-response analysis. Human glioma specimens were acquired on the basis of their spatial order or longitudinal pairing and were subjected to genomic analysis for identification of tumor-initiating (truncal) events. PDCs, patient-derived cells; 5-ALA, 5-aminolevulinic acid. (**b**) Somatic mutations, including single-nucleotide variants (SNVs) and small insertions/deletions, copy number alterations, and gene fusions, for 83 glioma multiregion or multisector longitudinal specimens from 30 patients are shown. 34 locally adjacent tumor fragments from 14 patients, 13 multifocal/multicentric (referred to as multiple) tissues from 5 patients, and a longitudinal pair, GBM14, with leptomeningeal seeding were collected from the Samsung Medical Center (SMC). We also curated 34 multisector longitudinal tumor exomes and/or RNA sequencing data from 10 patients in The Cancer Genome Atlas (TCGA) cohort[8]. All somatic mutations called by SAVI with an allele frequency of >5% are shown. For each gene, we calculated copy number (CN) on the basis of Excavator results. Clonal alterations were determined using ABSOLUTE with a cancer cell fraction of >80% (Online Methods). CNV, copy number variation.

sample, in 4 of 15 (26.7%) of the *EGFR*-amplified tumors, including 2 multiple cases (GBM5 and GBM9). Furthermore, *EGFR* mutations were shared by 3 of 7 (42.9%) cases, including one harboring disjoint alterations, which are different genomic alterations in the same gene (GBM7-I1: p.Leu62Arg and p.Arg108Lys; GBM7-I2: p.Ala289Val and p.Cys624Ser), suggesting that partial genetic information from a single tumor biopsy can be inconclusive in assessing the potential benefit from EGFR-targeted therapy (**Supplementary Fig. 1b**,**c**).

To understand the association between spatiotemporal architecture and genetic relevance, we calculated Nei's genetic distances among multisector samples from the same individual (Online Methods). Genetic diversity was greater in multiple tumors than in locally adjacent tissues ($q = 4.7 \times 10^{-5}$, Wilcoxon rank-sum test; **Fig. 2a**), in distant recurrences than in local recurrences ($q = 1.4 \times 10^{-5}$, Wilcoxon rank-sum test; **Fig. 2a**), and in long-term recurrence than in short-term recurrences ($q = 2.9 \times 10^{-3}$, Wilcoxon rank-sum test; **Fig. 2a**). Multinomial logistic regression was applied to classify multisector sample pairs on the basis of their genomic features (**Supplementary Note**). This analysis highlighted that tumors from distant regions or long-term recurrences, separated by surgical intervals exceeding 18 months, constitute a distinct evolutionary scenario in GBM (**Fig. 2b**). In colorectal tumors, a Big Bang model[20] shows that cells from different biopsies of the same tumor share clonal and subclonal variants (**Fig. 2c**, left). In accordance with this model, samples taken from locally adjacent tumors shared a large proportion of clonal and subclonal events (**Supplementary Fig. 2a**,**b**). In contrast, multiple tumors contained fewer shared (more private) clonal mutations when compared to local tumors ($q = 1.86 \times 10^{-3}$, Wilcoxon rank-sum test; **Supplementary Fig. 2a**). We corroborated this finding by computing statistics on the space of evolutionary trees (evolutionary moduli spaces; Online Methods and **Supplementary Fig. 3**). Locally adjacent tumors clustered near the tip of the space, indicating a higher shared mutation ratio than that of multiple tumors ($P = 1.27 \times 10^{-2}$, Wilcoxon rank-sum test). These results indicate that, in contrast to locally adjacent tumors, geographically separated multifocal tumors and/or long-term recurrent tumors are seeded from distinct clones, a phenomenon we call the 'multiverse model' (**Fig. 2c**, right). Unlike the Big Bang model[20], in the multiverse model, tumor samples that are derived from different tumor masses share very few genomic alterations, indicating that tumor clones are geographically segregated at an early stage of evolution and that each clone acquires distinct private alterations, leading to the construction of multiple universes.

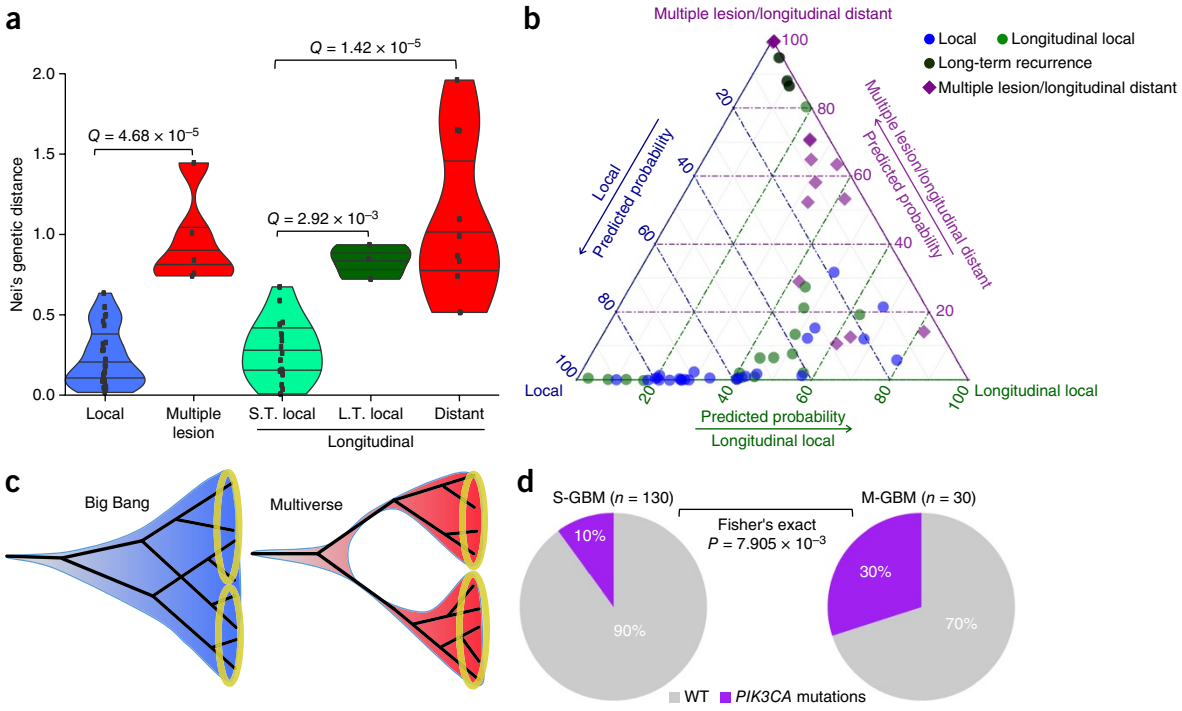

Figure 2 Comparison of genetic heterogeneity across glioma multisector and longitudinal samples. Patient samples were classified into five groups for comparative analyses: local, multiple lesion, S.T. (short-term) longitudinal local, L.T. (long-term) longitudinal local, and longitudinal distant. (**a**) Nei's genetic distances for each of the indicated groups are shown. *q* values were calculated by Wilcoxon rank-sum test and corrected for false discovery rate (FDR) using the Benjamini–Hochberg method. S.T. and L.T. local correspond to short-term (<18-month surgical interval) and long-term (≥18-month interval) recurrent tumors, respectively. The lines within each 'violin' represent the 25th, 50th, and 75th quantiles. $n = 25$ (local), 5 (multiple lesion), 17 (S.T. local), 3 (L.T. local), and 10 (distant). (**b**) Illustration of leave-one-out results from multinomial logistic regression. Each point represents one pair of samples, and the coordinates correspond to the probability that the pair is local, multiple lesion/longitudinal distant, or longitudinal local. L.T. recurrent samples were classified together with multiple lesion/longitudinal distant samples, indicating that they might follow the same evolutionary model. (**c**) Tumor evolution behind the Big Bang and multiverse models. The Big Bang model is represented as a mixture of tumor cells that share many clonal and subclonal alterations. The multiverse model is represented by a greater proportion of private events at a clonal level. (**d**) Pie charts show the frequencies of *PIK3CA* mutations in multifocal/multicentric glioblastomas (M-GBMs) (30%, 9/30) and solitary glioblastomas (S-GBMs) (10%, 13/130). WT, wild type. The *P* value was calculated using Fisher's exact test.

Next, we investigated the mutation profiles of GBMs with multifocal/multicentric lesions (M-GBMs) or solitary lesions (S-GBMs) in a total of 160 treatment-naive individuals from both the SMC and TCGA[17] cohorts (Online Methods, **Supplementary Figs. 4** and **5a**, and **Supplementary Table 3**). Notably, nonsynonymous mutations of *PIK3CA* were enriched in M-GBM (13/130 S-GBM and 9/30 M-GBM tumors; $P = 7.905 \times 10^{-3}$, Fisher's exact test; **Fig. 2d**). This conclusion remained the same in the *IDH1*-wild-type cohort (**Supplementary Fig. 5b**). *PIK3CA* induces multipotency of mammary tumors, suggesting that it has an associative role in tumor multiplicity[21]. Survival analysis indicated that both patients with M-GBM and *PIK3CA*-mutant patients had a worse prognosis than patients with S-GBM or wild-type *PIK3CA* ($P = 0.0151$ and 0.039, respectively, log-rank test; **Supplementary Fig. 5c**,**d**).

To further characterize the heterogeneity of expression profiles, we curated single-cell RNA–seq data from a total of seven different samples from three patients. Overall, expression-based cell subtypes were not clearly determined by location or time (**Fig. 3**), which is consistent with a previous report[10]. To make sure that this observation was not due to the limitations of this classification and to capture the transcriptional similarity among different cells, we used topological data analysis (Online Methods), a recently developed technique that summarizes and reduces the dimensionality of large data sets while retaining local high-dimensional structure[22,23].

GBM9 (**Fig. 3a–d**) consisted of samples from two initial tumors in the right and left frontal lobes and a recurrent tumor in the left frontal lobe that emerged after concurrent chemoradiotherapy (CCRT) and EGFR-targeted treatment (**Supplementary Note**). We found in bulk whole-exome sequencing (WES) and confirmed using ultra-deep sequencing (**Supplementary Table 4**) and single-cell analysis that cells from the recurrent tumor shared genomic and expression features with initial tumor cells from the left frontal lobe (**Fig. 3b** and **Supplementary Figs. 6–8**). Particularly, there were 61 somatic mutations shared by the left initial tumor and the recurrent tumor, while there were only 42 shared by the right initial tumor and the recurrent tumor. Single-cell transcriptome analysis showed *EGFR* expression predominantly in the right tumor mass, but not in the left initial and recurrent tumors (**Fig. 3c**). Different single cells harbored different *EGFR* alterations, implying that these alterations were late events during tumor evolution[21]. *PIK3CA* mutations were detected from single cells in all three samples, which is consistent with the bulk WES result that *PIK3CA* mutations are founder events (**Supplementary Figs. 1a** and **6a**). Our analysis also showed the presence of transcriptional heterogeneity within the individual samples. A subset of the left initial tumor cells was characterized by upregulation of mitotic genes that was not found in either the right or recurrent sections (**Fig. 3d**).

Additionally, we profiled *IDH1*-mutant tumor cells, which are distinguished by their 5-aminolevulinic acid (5-ALA) uptake

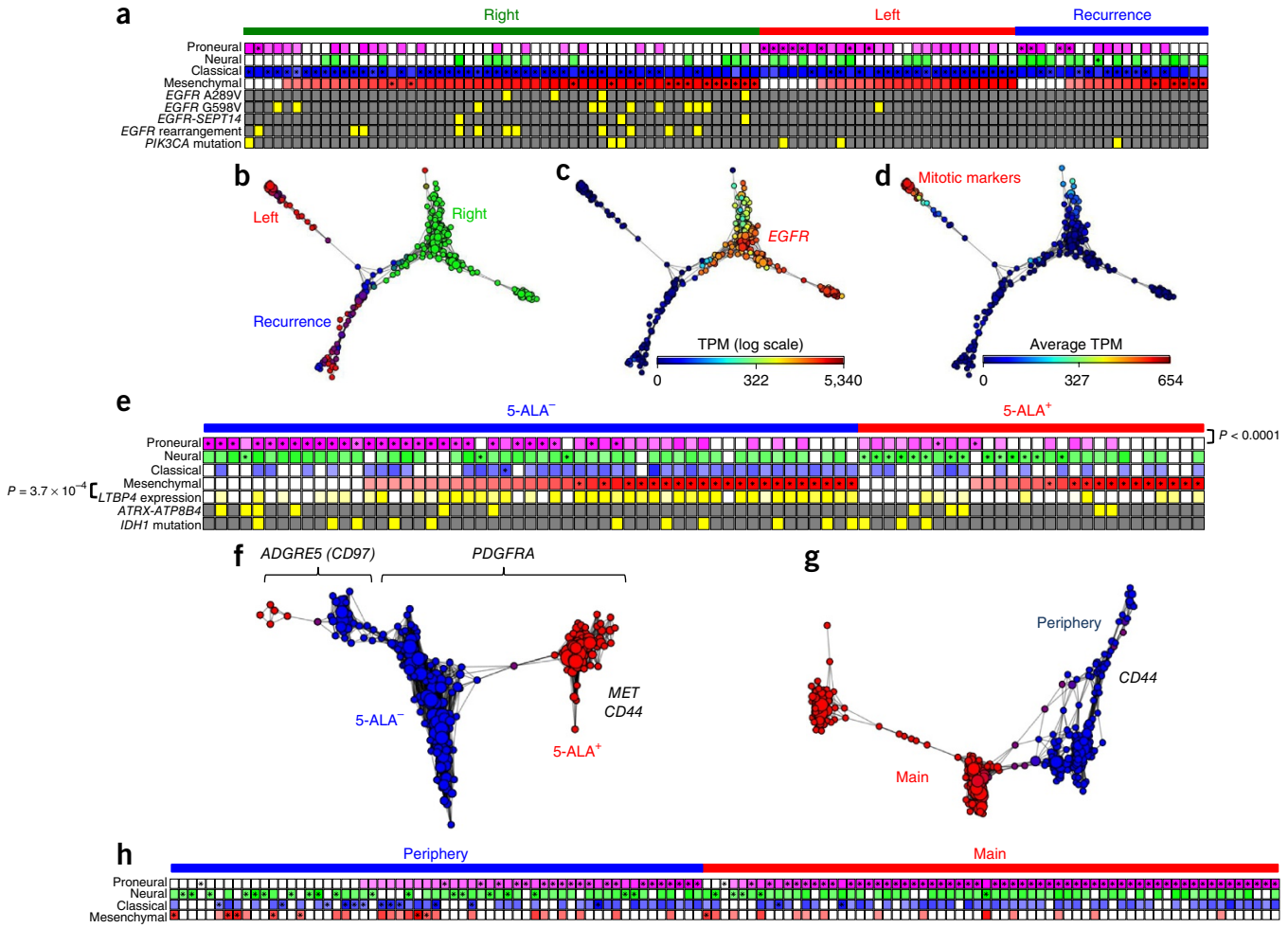

Figure 3 Single-cell transcriptomes from multiregion samples. (**a**) Expression profiles of individual tumor cells from three samples of GBM9 (left initial, right initial, and relapse) according to expression subtype[32]. For each cell, the subtype with the highest expression is marked with an asterisk. Several *EGFR* genomic alterations could be identified in the single-cell expression data (yellow), despite the abundance of missing data (gray). (**b**) Topological representation of the expression data of individual tumor cells from GBM9, labeled by sample of origin. Each node represents a set of cells with similar transcriptional profiles. A cell can appear in several nodes, and two nodes are connected by an edge if they have at least one cell in common. (**c**,**d**) Topological representations of GBM9 where tumor cells are labeled by expression of *EGFR* (**c**) and mitotic genes (**d**). TPM, transcripts per million. (**e**) Expression profiles of individual tumor cells from GBM10 (two samples: 5-ALA$^+$ and 5-ALA$^-$). The *P* value for comparison of the proneural and 5-ALA profiles was obtained using Fisher's exact test. The *P* value for comparison of the mesenchymal profile and *LTBP4* expression was calculated using Spearman's correlation. *ATRX* fusion was validated by RT–PCR assays (**Supplementary Fig. 7b,c**). (**f**,**g**) Topological representations of expression data for individual tumor cells from patients GBM10 (**f**) and GBM2 (**g**), where tumors cells are labeled by the sample of origin. (**h**) Expression profiles of individual tumor cells from patient GBM2 shown according to GBM expression subtype. Asterisks indicate the representative subtype identities of the corresponding cells.

pattern (populations stained for tumor cellularity[24]) (GBM10; **Fig. 3e**,**f** and **Supplementary Figs. 6** and **7**). Previous glioma studies suggested that a low pathologic grade is associated with a low rate of 5-ALA uptake[24,25]; however, genomic determinants for 5-ALA uptake remain elusive. Using single-cell transcriptome analysis, we found predominant enrichment of proneural cells in the 5-ALA$^-$ sample, supporting previous observations that GBM cells may evolve from proneural precursors ($P < 0.01$, Fisher's exact test; **Fig. 3e**)[26]. Also, we found enrichment of expression for several cell proliferation and migration markers in the 5-ALA$^+$ section, including *MET* and *CD44* (refs. 27–29). Notably, 5-ALA$^-$ tumors, which are considered to be less aggressive, are actually fully fledged tumors that harbor driver mutations and express markers for aggressive tumors (**Fig. 3e**,**f** and **Supplementary Figs. 6b**, **7**, **9a**, and **10**). Finally, we studied main tumor and resection margin samples from GBM2—a locally adjacent hypermutated case—and found distinct subpopulations of cells expressing mitotic cell markers and migration-associated genes, including *CD44* (**Fig. 3g** and **Supplementary Figs. 6**, **7**, **9**, and **10**)[28].

To investigate the influence of genetic heterogeneity on drug response, we isolated 28 PDCs from 11 individuals and screened 40 different cancer-related compounds (**Supplementary Table 5**)[4,30]. We found that Nei's genetic distance was associated with drug-response correlation ($P = 0.02$, Wilcoxon rank-sum test; **Fig. 4a**). Consistently, both distant and longitudinal samples showed significantly broader drug responses than local samples (**Fig. 4b**). We found that the PDCs from M-GBMs were more sensitive to PI3K–AKT–mTOR (PAM) pathway inhibitors than PDCs from solitary tumors were ($P = 1.872 \times 10^{-6}$, Wilcoxon rank-sum test; **Fig. 4c** and **Supplementary Fig. 11**). This indicates that PAM inhibitors could provide a clinical benefit for patients with M-GBM. In addition, we observed that PDCs from recurrent GBMs were more resistant to EGFR inhibitors than

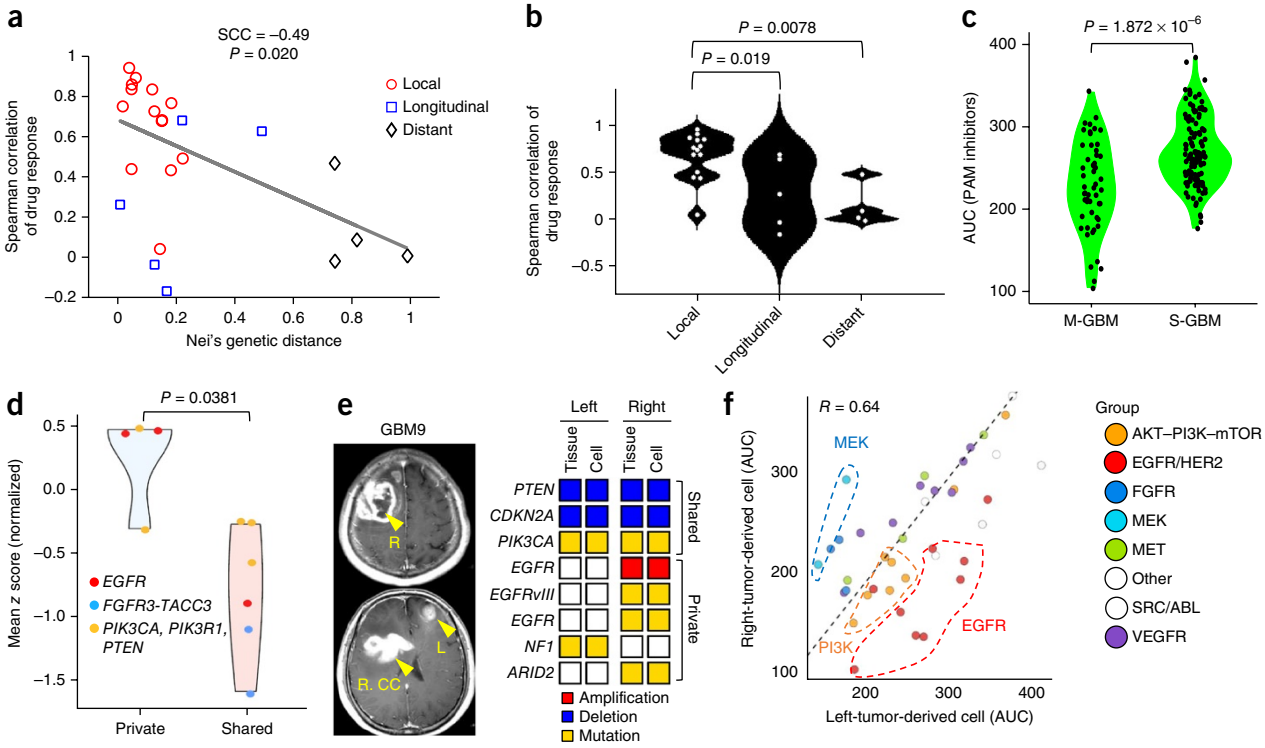

Figure 4 Chemical screening of multiregion PDCs. (**a**) PDCs were treated with 40 chemical agents targeting oncogenic signaling pathways in dilution series from 20 μM to 4.88 nM. The *x* axis shows Nei's genetic distances between fragments from the same patient, and the *y* axis shows the Spearman's correlation coefficient (SCC) of corresponding fragments based on drug sensitivities as measured by area under the curve (AUC). (**b**) Violin plots for the SCC of drug responses for the groups described in **a**. Each white dot represents a tumor sample. $n = 14$ (local), 5 (longitudinal), and 4 (distant). (**c**) Mean values of the AUCs for six PI3K–AKT–mTOR (PAM) inhibitors (BEZ235, BKM120, BYL719, AZD5363, AZD2014, and everolimus) for PDCs isolated from M-GBMs ($n = 9$) or S-GBMs ($n = 22$). Each black dot represents a mean. (**d**) The normalized *z* score for each PDC obtained from the AUCs for the indicated drug classes used in **a** and **b** was plotted when the corresponding tissues harbored associated genetic alterations, designated as 'shared' or 'private'. A genetic alteration was determined to be private when the drug-response-associated genetic alteration was private, and vice versa for the shared group. For example, *PTEN* mutations for P13K–AKT–mTOR pathway inhibitors were private, meaning they were exclusive to only one or two regions of the multisector sample. (**e**) Preoperative T1-weighted contrast-enhanced magnetic resonance images (MRIs) and key genomic alterations found in the corresponding tumors and their derivative cells for a patient with multicentric disease (GBM9). "R" indicates right-side tumors that encompassed the right frontal lobe and corpus callosum (CC). "L" indicates the left frontal lobe tumor. The preoperative MRIs show a multifocal infiltrative lesion in both the frontal lobes and CC. (**f**) Scatterplot of the AUCs for 40 cancer-targeting compounds on GBM9 PDCs that were derived from the left- and right-side tumors in the frontal lobes. Each dot represents the AUC of a given drug. The *R* value was obtained as Pearson's correlation coefficient. All *P* values in this figure were obtained using Wilcoxon rank-sum tests.

the initial tumor cells were ($P = 2.9 \times 10^{-4}$, Wilcoxon rank-sum test; **Supplementary Figs. 12** and **13**).

We hypothesized that clonal alterations found in all multisector samples (truncal alterations) represent better molecular targets than those found in only a subset of multisector samples (private alterations). In agreement with this truncal-target hypothesis (**Supplementary Video 1**), multisector PDCs were more sensitive to drugs that target shared alterations than to drugs that target private alterations ($P = 0.0381$, Wilcoxon rank-sum test; Online Methods, **Fig. 4d** and **Supplementary Figs. 14–16**). The multiverse model implies that the extensive genetic diversity of multiple tumors presents a special challenge. Accordingly, GBM9 showed a divergent genetic profile and a highly heterogeneous drug response (**Fig. 4e**,**f**). PDCs from the right-side tumor were highly sensitive to EGFR inhibitors, but not to MEK inhibitors, and vice versa for the left-side tumor. However, inhibitors of the PAM pathway were ubiquitously effective, which is consistent with our hypothesis that targeting the PAM signaling pathway could be a potent option to treat M-GBMs (**Fig. 4c**,**f**). Yet not all truncal alterations can serve as drug targets. For example, gatekeeper genes, which are necessary for tumor initiation but are no longer required for tumor maintenance, are not good candidates[31]. Although targeting subclonal mutations shows a limited effect, patients might still benefit from the elimination of a subclone that has a bystander effect on surrounding cells.

In conclusion, on the basis of comprehensive bulk and single-cell analyses, we have proposed a multiverse model to interpret the evolution of multiple GBMs. We showed that M-GBMs are more genetically diverse than locally adjacent tumors and that genetic similarity between multiregion samples is associated with consistent drug response. Specifically, we found an enrichment of *PIK3CA* mutations in M-GBMs, and we found that inhibitors of the PAM pathway are more effective in PDCs from this cohort. These findings support the truncal-target hypothesis, which states that truncal mutations can inform more effective therapies.

URLs. R nnet package, multinom function, https://CRAN.R-project.org/package=nnet; Mapper algorithm, http://danifold.net/mapper; http://github.com/MLWave/kepler-mapper.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by a grant of the Korea Health Technology R&D project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI14C3418). R.R. acknowledges funding from the NIH (U54 CA193313, R01 CA185486, R01 CA179044). J.W. is supported by Precision Medicine Fellowship (UL1 TR000040). E.L. is supported by NIH (F99 CA212478) and Cancer Biology Training Program (T32 CA09503).

AUTHOR CONTRIBUTIONS

J.-K.L., J.W., J.K.S., and E.L. are co-first authors. J.-K.L., J.W., J.K.S., and E.L. performed the majority of experiments and analyses. W.-Y.P., H.-O.L., K.-T.K., P.G.C., and P.v.N. performed experiments and analyses for the single-cell transcriptome. D.S.R., Z.L., A.B., A.C., Y.J.S., and S.S. conducted several experiments and analyses. D.-H.N., D.-S.K., H.J.S., C.-K.P., and J.-I.L. provided surgical specimens. S.W.J., S.W.C., and J.K. helped interpret clinical data. W.-Y.P., I.-H.L., Y. J.S., J.-M.O., and H.J.K. organized and processed specimens and genome data. J.-K.L., J.W., J.K.S., E.L., and P.G.C. wrote the manuscript with feedback from R.R., A.I., and D.-H.N. D.-H.N. and R.R. designed and supervised the entire project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hamburg, M.A. & Collins, F.S. The path to personalized medicine. *N. Engl. J. Med.* **363**, 301–304 (2010).
2. Brennan, C.W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
3. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
4. Cloughesy, T.F., Cavenee, W.K. & Mischel, P.S. Glioblastoma: from molecular pathology to targeted treatment. *Annu. Rev. Pathol.* **9**, 1–25 (2014).
5. Frattini, V. *et al.* The integrated landscape of driver genomic alterations in glioblastoma. *Nat. Genet.* **45**, 1141–1149 (2013).
6. Snuderl, M. *et al.* Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810–817 (2011).
7. Szerlip, N.J. *et al.* Intratumoral heterogeneity of receptor tyrosine kinases *EGFR* and *PDGFRA* amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc. Natl. Acad. Sci. USA* **109**, 3041–3046 (2012).
8. 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).
9. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. USA* **110**, 4009–4014 (2013).
10. Patel, A.P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
11. Kumar, A. *et al.* Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. *Genome Biol.* **15**, 530 (2014).
12. Johnson, B.E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, 189–193 (2014).
13. Yates, L.R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
14. Wang, J. *et al.* Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776 (2016).
15. Ohka, F., Natsume, A. & Wakabayashi, T. Current trends in targeted therapies for glioblastoma multiforme. *Neurol. Res. Int.* **2012**, 878425 (2012).
16. Kim, J. *et al.* Spatiotemporal evolution of the primary glioblastoma genome. *Cancer Cell* **28**, 318–328 (2015).
17. Liu, Q. *et al.* Genetic, epigenetic, and molecular landscapes of multifocal and multicentric glioblastoma. *Acta Neuropathol.* **130**, 587–597 (2015).
18. Singh, D. *et al.* Transforming fusions of *FGFR* and *TACC* genes in human glioblastoma. *Science* **337**, 1231–1235 (2012).
19. Hartmann, C., Bartels, G., Gehlhaar, C., Holtkamp, N. & von Deimling, A. *PIK3CA* mutations in glioblastoma multiforme. *Acta Neuropathol.* **109**, 639–642 (2005).
20. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* **47**, 209–216 (2015).
21. Koren, S. *et al.* $PIK3CA_{H1047R}$ induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114–118 (2015).
22. Carlsson, G. Topology and data. *Bull. Amer. Math. Soc.* **46**, 255–308 (2009).
23. Singh, G., Mémoli, F. & Carlsson, G.E. Topological methods for the analysis of high dimensional data sets and 3D object recognition. in *SPBG* (eds. Botsch, M. & Pajarola, R.) 91–100 (Citeseer, 2007).
24. Moiyadi, A., Syed, P. & Srivastava, S. Fluorescence-guided surgery of malignant gliomas based on 5-aminolevulinic acid: paradigm shifts but not a panacea. *Nat. Rev. Cancer* **14**, 146 (2014).
25. Stummer, W. *et al.* Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *Lancet Oncol.* **7**, 392–401 (2006).
26. Ozawa, T. *et al.* Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* **26**, 288–300 (2014).
27. Merzak, A., Koocheckpour, S. & Pilkington, G.J. CD44 mediates human glioma cell adhesion and invasion *in vitro*. *Cancer Res.* **54**, 3988–3992 (1994).
28. Yoshida, T., Matsuda, Y., Naito, Z. & Ishiwata, T. CD44 in human glioma correlates with histopathological grade and cell migration. *Pathol. Int.* **62**, 463–470 (2012).
29. Joo, K.M. *et al.* MET signaling regulates glioblastoma stem cells. *Cancer Res.* **72**, 3828–3838 (2012).
30. Lemmon, M.A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
31. Kinzler, K.W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
32. 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).

ONLINE METHODS

Glioma specimens and their derivative cells. After receiving informed consents, glioma specimens and clinical records were obtained from patients undergoing surgery at Samsung Medical Center (SMC) or Seoul National University Hospital (SNUH) in accordance with its institutional review board (IRB file no. 2010-04-004). Surgical samples measuring approximately 5 × 5 × 5 mm^3 were snap-frozen using liquid nitrogen for genomic analysis. We also curated whole-exome and/or RNA sequencing of 33 multisector specimens from 10 glioblastoma (GBM) patients in The Cancer Genome Atlas (TCGA) cohort[8] and 22 previously reported GBM longitudinal pairs[14,16]. To investigate the genomic characteristics of solitary and multifocal/multicentric GBMs, we curated exome sequencing data for 83 and 77 tumors with matched normal DNA from the SMC and TCGA cohorts[8,17], respectively. Portions of the surgical samples were enzymatically dissociated into single cells, following the procedures reported previously with modification of immune cell depletion[33]. Tumor cells were cultured in neurobasal medium with N2 and B27 supplements (0.5× each, half of the suggested working concentration; Invitrogen), human recombinant basic fibroblast growth factor (bFGF), and epidermal growth factor (EGF; 20 ng/ml each; R&D Systems). The patient-derived cells (PDCs) used here had shown no obvious contamination of mycoplasma.

Radiological evaluation. Both T1-weighted contrast enhancement (T1CE) and fluid-attenuated inversion recovery (FLAIR)/T2 axial images of 160 treatment-naive GBMs (83 and 77 tumors from the SMC and TCGA cohorts, respectively) were reviewed. Magnetic resonance images (MRIs) of tumors from the TCGA cohort have been obtained from The Cancer Imaging Archive (TCIA) website[17,34,35]. We excluded cases with any evidence of prior neurosurgical intervention except biopsy, lack of treatment history, or loss of T1CE or FLAIR/T2 images. To distinguish the multifocal/multicentric GBMs (M-GBMs) from solitary ones (S-GBMs), we adapted annotations from the VASARI feature set for human glioma[36]. According to the VASARI feature set, m-GBMs are defined as having at least one region of tumor, either enhancing or nonenhancing, that is not contiguous with the main lesion and is outside of the region of signal abnormality (edema) surrounding the main mass[17,37,38]. When a FLAIR/T2 high-signal-intensity lesion resides outside of the T1CR lesion, it is considered a separate tumor focus and is counted as a multifocal tumor in our study[39,40]. In contrast, tumors that present separate contrast-enhancement lesions within the FLAIR/T2 high-signal-intensity background are considered as solitary ones.

Whole-exome sequencing. An Agilent SureSelect kit was used to capture the exonic DNA fragments. An Illumina HiSeq 2000 instrument was used for sequencing and generated 2 × 101-bp paired-end reads.

Somatic mutation. The sequenced reads in the FASTQ files were aligned to the human genome assembly (hg19) using Burrows–Wheeler aligner version 0.6.2. The initial alignment BAM files were subjected to conventional preprocessing before mutation calling: sorting, removing duplicated reads, locally realigning reads around potential small indels, and recalibrating base quality scores using SAMtools, Picard version 1.73, and Genome Analysis Toolkit (GATK) version 2.5.2. We used MuTect (version 1.1.4) and Somatic IndelDetector (GATK version 2.2) to make high-confidence predictions on somatic mutations from the neoplastic and non-neoplastic tissue pairs. Variant Effect Predictor (VEP) version 73 was used to annotate the called somatic mutations. Additionally, we ran SAVI (Statistical Variant Identification) software to call somatic variants and indels in order to for refine the mutation calls from the above pipeline.

Copy number. Excavator was used to generate estimated copy number alterations in a tumor specimen in comparison with its matching, non-neoplastic part. For each gene, we calculated copy number = 2^{x+1}, where x is the segmentation mean from Excavator, which is defined as the $\log_2$ (fold change) in the tumor divided by the normal sample. The gene was labeled as 'amplified' when the copy number was 3 and 'deleted' when it was ≤1.

Cancer cell fractions and clonality. We ran ABSOLUTE[41] using input of genomic variants and copy number data to infer sample purity and cancer cell fractions (CCFs) and removed those that had <20% purity. We considered mutations as clonal if they were indicated as clonal in ABSOLUTE and had a CCF of at least 80% or if they had a CCF of 100% and were not marked as clonal or subclonal. The ABSOLUTE CCF estimates with regard to hypermutated samples appeared disproportionately subclonal in sample GBM18 initial and in TCGA-14-1402 second recurrence; we reasoned that the large mutational load might skew estimates. In hypermutated samples, treatment-associated mutation coupled with defects in mismatch repair is deemed largely responsible for a majority of observed mutations. Therefore, mutations having CCFs greater than or equal to the maximum mismatch repair CCF were marked clonal in these two samples.

If a mutation was found to be clonal in all sectors of a patient's tumor, it was inferred to be clonal throughout the entire tumor. We investigated the number of sequenced tumor sectors or cores needed to obtain a reasonable false discovery rate (FDR) for this inference of clonality. We analyzed glioma patient LGG174, published recently by Suzuki *et al.*, where nine sectors from different locations of the same tumor mass were sequenced[42]. On the basis of Figure 7b from that paper[42], 13 mutations have a high tumor cell fraction (>60%) shared by all samples. To relate the number of sectors that were sequenced to the number of mutations deemed to be clonal tumor-wide, we exhausted all possible subsampling strategies (number of cores k = 1, 2, …, 9) and calculated the reported clonal mutations based on k cores. For example, if there are two cores (k = 2), there are $C_2^9 = 36$ potential sampling strategies. We found that 22 of 36 sampling strategies contained no false discoveries in identifying clonal mutation. For each value of k, we calculated the FDR (**Supplementary Fig. 17**). Almost 90% of clonal mutations identified by two-core sequencing are true clonal mutations, and over 95% identified by three-core sequencing are true clonal mutations.

Nei's genetic distances. Nei's genetic distance is used in population genetics to assess the similarity between populations, taking into account heterogeneity within populations. Samples containing the same spatial or longitudinal category (local, 5-ALA, multiple lesion, longitudinal local, longitudinal distant) were retained for statistical comparisons. We calculated Nei's genetic distance of CCF for each patient's sample as follows. Let x be all CCFs of sample 1 and y be all CCFs of sample 2.

$$D = -\log \frac{\sum(x_i y_i + (1-x_i)(1-y_i))}{\sqrt{(\sum x_i^2 + (1-x_i)^2)(\sum y_i^2 + 1 - y_i)^2)}}$$

The multiverse model of tumor evolution. We found an increased Nei's genetic distance in multifocal/multicentric biopsies when compared with those that were locally adjacent. In addition, private clonal mutations appear frequently in multisectional and distant longitudinal samples, but are infrequent in local samples (**Supplementary Fig. 2a**). This spurred a hypothesis that specific early event(s) can give rise to distinct mutational profiles in spatially separated tumors (**Fig. 2a**). These differences in mutational load suggested that distinct tumor profiles might arise in separate 'universes' of clones rather than from one large growth period followed by diversification.

For each somatic mutation, we recorded the clonal status as determined by ABSOLUTE and whether the mutation is shared or private, or the clonal status changed between biopsies. Mutations are then classified into five patterns with respect to every available pair of a patient's samples. The mutational classes were labeled as the following: CC (clonal–clonal), CS (clonal–subclonal), SS (subclonal–subclonal), CX (clonal–absent), or SX (subclonal–absent). The order of the sample pair was not important: a mutation that was clonal in one sample and subclonal in the other was marked "CS," regardless of sample identity.

These mutational classifications were used to predict whether the spatiotemporal configuration of a sample pair fell into one of three groups: locally adjacent, local longitudinal, or multisectional/distant longitudinal. The fractions of mutations in a sample pair that fit each of the five patterns were used as features in a multinomial logistic regression. Predictions were then made using leave-one-out cross-validation.

Mutational pairs plotted on the simplex allowed us to visually separate multisectional/distant longitudinal, locally adjacent, or local longitudinal sections in agreement with most of our MRI classifications. The simplex

doi:10.1038/ng.3806

axes represent the predicted probabilities of outcomes for each observation. The sample layout contained three local longitudinal outliers closest to the multisectional point of the simplex. The time interval between surgeries for the three pairs was 18 months or more. Moreover, their Nei distances were significantly different from all other sections ($P = 0.01652$). We labeled all samples exceeding surgical intervals of 18 months as long-term recurrence and colored them in dark green. Analysis was performed in the R computing environment using the multinom function from the nnet package (see "Data availability").

Isolation of single cells and RNA sequencing. We adopted the C1TM Single-Cell Auto Prep System (Fluidigm) with the SMARTer kit (Clontech) to generate cDNAs from single cells. Cells were captured as a single isolate on a C1 chip (17–25 μm), as determined from bright-field images obtained under 100× magnification using an Axiovert200 inverted microscope (Carl Zeiss). RNAs from pooled samples were also processed using the SMARTer kit with 10 ng of starting materials. Libraries were generated using the Nextera XT DNA Sample Prep Kit (Illumina) and sequenced on the HiSeq 2500 using the 100-bp paired-end mode of the TruSeq Rapid PE Cluster kit and TruSeq Rapid SBS kit. Before mapping RNA sequencing reads to the reference, reads were filtered out at Q33 by using Trimmomatic-0.30. Transcripts per million (TPM) values were calculated from each single cell (as if they were different samples) using RSEM (version 1.2.25)[43] and are expressed as $\log_2$ (1 + TPM).

Gene fusion detection. Chimerascan was applied to generate a list of candidate gene fusions[44]. For bulk sequencing, only previously reported in-frame, highly expressed fusions, such as *FGFR3-TACC3* (ref. 18), *MGMT* fusion[14], *EGFR-SEPT14* (ref. 5) and *ATRX* fusion, were considered in this manuscript. For single-cell fusion analysis, if a fusion was highly expressed and independently detected in multiple cells, the fusion was reported.

Expression-based subtype determination. Gene expression was measured by RSEM and then $\log_2$ transformed. To determine the expression-based subtype of GBM cells, we first calculated *z* scores for gene expression data across samples and then applied ssGSEA (version gsea2-2.2.1) to the normalized expression profile. For each cell, we ranked all genes on the basis of their expression values to create a .rnk file as input for the software GseaPreranked. An enrichment score was computed for all four subtypes initially defined in Verhaak *et al.*[32]. The subtype with the maximal enrichment score was used as the representative subtype for each cell.

Topological data analysis using single-cell transcriptomes. We filtered out normal cells on the basis of their expression profile. To that end, we considered the expression signatures of normal oligodendrocytes, neurons, astrocytes[45], microglia[46], endothelial cells[47], T cells[48], and other immune cells[48,49], and we used a Gaussian mixture model to classify individual cells according to their expression profile. 94/133, 82/85, and 90/137 cells for GBM9, GBM10, and GBM2, respectively, were classified as tumor cells. After normalizing the gene expression level by dividing by the total number of reads in each cell to eliminate potential bias caused by batch effect, we built topological representations of these single-cell data using the Mapper algorithm[23], as implemented by Ayasdi. Open-source implementations of this algorithm are also available (see "Data availability"). We used the first two components of multidimensional scaling (MDS) as auxiliary functions for the algorithm. The output of Mapper is a low-dimensional network representation of the data, where nodes represent sets of cells with similar global transcriptional profiles (as measured by correlation of the expression levels of the 2,000 genes with the highest variance across each patient). We identified individual genes that had an expression pattern localized in the network and used these to determine the subclonal structure of the samples at the level of expression.

PDC-based chemical screening and analysis. PDCs grown in serum-free medium were seeded in 384-well plates at a density of 500 cells per well in duplicate or triplicate for each treatment. The drug panel consisted of 40 anticancer agents (Selleckchem) targeting oncogenic signals. Two hours after plating, PDCs were treated with drugs in a fourfold and seven-point serial dilution series from 20 μM to 4.88 nM using the Janus Automated Workstation (PerkinElmer). After 6 d of incubation at 37 °C in a 5% CO_2 humidified incubator, cell viability was analyzed using an ATP-monitoring system based on firefly luciferase (ATPLit 1step, PerkinElmer). The number of viable cells was estimated using the EnVision Multilabel Reader (PerkinElmer). DMSO was also included as a control in each plate. Controls were used for the calculation of relative cell viability for each plate, and normalization was performed on a per-plate basis. Dose-response curve (DRC) fitting was performed using GraphPad Prism 5 (GraphPad) and evaluated by measuring the area under the curve (AUC) of the DRC. After normalization, best-fit lines were determined and the AUC value of each curve was calculated using GraphPad Prism, ignoring regions defined by fewer than two peaks.

Cell viability was determined via calculating AUC values of DRCs with exclusion of nonconvergent fits[50].

Moduli space analysis. To illustrate the evolution histories of tumors in GBM patients, we applied moduli space analysis[51] in local and multiple groups of patients. Multiregion pairs were compared to calculate the number of shared and private mutations. In this analysis, clonal mutations were separated on the basis of their allele frequencies. Sector pairs were put in the left sphere on the basis of the number of shared and private mutations with high-allele-frequency mutations (>20%), whereas the same number of pairs were put in the right sphere on the basis of mutations with low-allele-frequency mutations (<20%). The same analysis was repeated using CCF instead of mutational allele fraction.

Immunohistochemistry. Tissue specimens were fixed by formalin and embedded in paraffin. Paraffin-embedded sections were treated with 0.3% hydrogen peroxide to block endogenous peroxidase activity, and antigens were retrieved by heating sections in 10 mM sodium citrate (pH 6.0) at 95 °C for 30 min. Sections were incubated with primary antibodies overnight at 4 °C, biotinylated secondary antibodies for 1 h at room temperature, and avidin–biotin complex for 1 h at room temperature.

Protein blotting. GBM PDCs were washed with cold PBS and harvested in lysis buffer (150 mM sodium chloride, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl, and 2 mM EDTA), and a protease and phosphatase inhibitor cocktail was added (Thermo Scientific). Insoluble materials were removed by centrifugation at 13,500*g* for 15 min at 4 °C. Proteins were separated by SDS–PAGE. Immunoblotting was performed using antibodies against indicated proteins.

Limiting dilution assays. GBM PDCs were dissociated into single-cell suspensions and then plated into 96-well plates at 1–250 cells per well. Cells were incubated at 37 °C for 1–2 weeks. At the time of quantification, each well was examined for formation of neurosphere-like cell aggregates. Statistical significance was evaluated using extreme limiting dilution analysis (ELDA; Walter+Eliza Hall Bioinformatics).

Gene fusion validation. Validation of gene fusion transcripts were performed by RT–PCR assays. Total RNA was extracted from tissues by AllPrep DNA/RNA Mini Kit according to the manufacturer's instructions (Qiagen). Total RNA (1 μg) was reverse transcribed to synthesize template cDNA by random hexamers using the SuperScript III First-Strand System (Life Technologies), and 20 μl of synthesized cDNA was diluted fivefold with DEPC-treated water. For RT–PCR, EzWay Taq PCR MasterMix (Komabiotech, Korea) and 5 μl of synthesized cDNA template were used. Thermal cycling was carried out under the following conditions: incubation at 95 °C for 1 min followed by 30 cycles of 30 s at 95 °C, 30 s at 56 °C, and 30 s at 72 °C. The primer pairs used in this experiment were designed to generate the amplification product, including the breakpoints of the fusion genes. PCR products were analyzed by agarose gel electrophoresis. The primers are summarized in **Supplementary Table 6**.

Data availability. All sequenced data have been deposited in the European Genome-phenome Archive (EGA) with accession code EGAS00001001880 (RNA–seq and WES data).

33. Lee, J. *et al.* Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* **9**, 391–403 (2006).
34. Clark, K. *et al.* The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* **26**, 1045–1057 (2013).
35. Wangaryattawanich, P. *et al.* Multicenter imaging outcomes study of The Cancer Genome Atlas glioblastoma patient cohort: imaging predictors of overall and progression-free survival. *Neuro-oncol.* **17**, 1525–1537 (2015).
36. Rios Velazquez, E. *et al.* Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. *Sci. Rep.* **5**, 16822 (2015).
37. Giannopoulos, S. & Kyritsis, A.P. Diagnosis and management of multifocal gliomas. *Oncology* **79**, 306–312 (2010).
38. Kyritsis, A.P., Levin, V.A., Yung, W.K. & Leeds, N.E. Imaging patterns of multifocal gliomas. *Eur. J. Radiol.* **16**, 163–170 (1993).
39. Zinn, P.O. *et al.* Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLoS One* **6**, e25451 (2011).
40. Li, Y.M., Suki, D., Hess, K. & Sawaya, R. The influence of maximum safe resection of glioblastoma on survival in 1229 patients: can we do better than gross-total resection? *J. Neurosurg.* **124**, 977–988 (2016).
41. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
42. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468 (2015).
43. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA–Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
44. Iyer, M.K., Chinnaiyan, A.M. & Maher, C.A. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903–2904 (2011).
45. Cahoy, J.D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
46. Butovsky, O. *et al.* Identification of a unique TGF-β-dependent molecular and functional signature in microglia. *Nat. Neurosci.* **17**, 131–143 (2014).
47. Bhasin, M. *et al.* Bioinformatic identification and characterization of human endothelial cell–restricted genes. *BMC Genomics* **11**, 342 (2010).
48. Chtanova, T. *et al.* Identification of T cell–restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. *J. Immunol.* **175**, 7837–7847 (2005).
49. Abbas, A.R. *et al.* Immune response *in silico* (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6**, 319–331 (2005).
50. Huang, S. & Pang, L. Comparing statistical methods for quantifying drug sensitivity based on *in vitro* dose-response assays. *Assay Drug Dev. Technol.* **10**, 88–96 (2012).
51. Zairis, S. *et al.* Moduli spaces of phylogenetic trees describing tumor evolutionary patterns. *Lect. Notes Comput. Sci.* **8609**, 528–539 (2014).

doi:10.1038/ng.3806