HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 October 04.

Published in final edited form as:

Nat Genet. 2017 April ; 49(4): 594–599. doi:10.1038/ng.3806.

Spatiotemporal genomic architecture informs precision oncology in glioblastoma

Jin-Ku Lee[1,2,#], **Jiguang Wang**[9,10,11,#], **Jason K. Sa**[1,2,4,#], **Erik Ladewig**[9,10,#], **Hae-Ock Lee**[3], **In-Hee Lee**[1,2], **Hyun Ju Kang**[1,2], **Daniel S. Rosenbloom**[9,10], **Pablo G. Camara**[9,10], **Zhaoqi Liu**[9,10], **Patrick van Nieuwenhuizen**[9,10], **Sang Won Jung**[1,2,4], **Seung Won Choi**[1,2,4], **Junhyung Kim**[1,2], **Andrew Chen**[9,10], **Kyu-Tae Kim**[3], **Sang Shin**[1,2,4], **Yun Jee Seo**[1,2], **Jin-Mi Oh**[1,2], **Yong Jae Shin**[1,2,3], **Chul-Kee Park**[5], **Doo-Sik Kong**[2], **Ho Jun Seol**[2], **Andrew Blumberg**[12], **Jung-Il Lee**[2], **Antonio Iavarone**[6,7,8], **Woong-Yang Park**[3,4,*], **Raul Rabadan**[9,10,*], and **Do-Hyun Nam**[1,2,4,*]

[1]Institute for Refractory Cancer Research, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

[2]Department of Neurosurgery, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

[3]Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

[4]Department of Health Science & Technology, Samsung Advanced Institute for Health Science & Technology, Sungkyunkwan University, Seoul 06351, Korea

[5]Department of Neurosurgery, College of Medicine, Seoul National University and Seoul National University Hospital, Seoul 03080, Korea

[6]Institute for Cancer Genetics, Columbia University, New York, NY10032

[7]Department of Neurology, Columbia University, New York, NY10032

[8]Department of Pathology, Columbia University, New York, NY10032

[9]Department of Systems Biology, Columbia University, New York, NY10032

[10]Department of Biomedical Informatics, Columbia University, New York, NY10032

[*]Correspondences: rr2579@cumc.columbia.edu (R. R.), woongyang@skku.edu (W.-Y. P.) and nsnam@skku.edu (D.-H. N.).
[#]These authors contributed equally to this work.

Disclosure of Potential Conflicts of Interest

The authors declare no competing financial interests.

Data Availability Statement

The European Genome-phenome Archive (EGA): Coordinates have been deposited with accession code EGAS00001001880 (RNA-seq and WES data).

Author Contributions

J.-K. Lee, J. Wang, J. K. Sa and E. Ladewig are co-first authors. J.-K. Lee, J. Wang, J. K. Sa and E. Ladewig performed the majority of experiments and analyses. W.-Y. Park, H.-O. Lee, K.-T. Kim, P. G. Camara and P. Nieuwenhuizen performed experiments and analyses for single cell transcriptome. D. Rosenbloom, Z. Liu, A. Blumberg, A. Chen, Y. J. Shin and S. Shin conducted several experiments and analyses. D.-H. Nam, D.-S. Kong, H. J. Seol and J.-I. Lee provided surgical specimens. S. W. Jung, S. W. Choi and J. Kim helped interpretation of clinical data. W.-Y. Park, I.-H. Lee, Y. J. Seo, J.-M. Oh, and H. J. Kang organized and processed specimens and genome data. J.-K. Lee, J. Wang, J. K. Sa, E. Ladewig, and P. G. Camara wrote the manuscript with the feedback from R. Rabadan, A. Iavarone and D.-H. Nam. D.-H. Nam and R. Rabadan designed and supervised the entire project.

[11]Division of Life Science and Division of Biomedical Engineering, Hong Kong University of Science and Technology, Hong Kong

[12]Department of Mathematics, University of Texas, Austin, TX78712

Abstract

Precision medicine in cancer proposes that genomic characterization of tumors can inform personalized targeted therapies[1–5]. This proposition, however, is complicated by spatial and temporal heterogeneity[6–14]. Here we study genomic and expression profiles across 127 multi-sector or longitudinal specimens from 52 glioblastoma (GBM) patients. Using bulk and single-cell data, we find that samples from the same tumor mass share genomic and expression signatures, while 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 to genetic similarity, and multifocal tumors enriched with *PIK3CA* mutations have a heterogeneous drug response pattern. Importantly, we show that targeting truncal events is more efficacious in reducing tumor burden. In summary, this work demonstrates that evolutionary inference from integrated genomic analysis in multi-sector biopsies can inform targeted therapeutic interventions for GBM patients.

Main Text

In glioblastoma (GBM) most clinical trials on targeted therapy 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, there is little understanding which strategies can effectively use genomic data to inform targeted therapies. To identify such strategies, we analyzed somatic variants in 127 multi-region or longitudinal tumor specimens from 52 glioma patients: 42 from Samsung Medical Center (SMC) Seoul, South Korea, and 10 from The Cancer Genome Atlas (TCGA) GBM cohort[8,14,16] (Supplementary Table 1). Additionally, we analyzed transcriptomes of 83 tumor specimens from 41 patients (bulk) and 305 single cells from seven samples of three patients. Tumors were classified into three distinct groups according to spatial and temporal tissue acquisition: from the same location and time (locally adjacent), from different locations at the same time (multifocal/multicentric, referred as multiple)[17], and from local and distant recurrences at different times (longitudinal local or distant, respectively) (Fig. 1a).

We inferred clonal and subclonal alterations from cancer cell fractions in multiple sectors (Methods, Supplementary Table 2). The average mutation rate was 2.2 mut/Mb for non-hypermutated samples, consistent with previous studies[2,14]. *IDH1* R132 mutations were clonal across all the regions in IDH mutant tumors[2,14] (6/6). *PIK3CA* mutations were always clonal and shared in all sectors (5/5), consistent with our previous longitudinal analysis from tumor evolution directed graphs (TEDG) that *PIK3CA* mutations are early events[14] (Fig. 1b, Supplementary Fig. 1a–b). Furthermore, *FGFR3-TACC3* fusions were highly expressed in all regions from two patients[18]. These somatic variants that are shared in all tumor regions represent promising therapeutic targets[14,19], as they reflect truncal alterations suspected to be present among all tumor cells. In contrast, *PTEN* alterations including copy number

deletions and mutations were shared in 10/20 (50%) and 5/7 (71.4%) tumor sectors, respectively. Likewise, *EGFR* amplifications were observed as private events in 4/15 (26.7%) of *EGFR*-amplified tumors, including two multiple cases (GBM5 and GBM9). Furthermore, *EGFR* mutations were shared in 3/7 (42.9%) cases, including one harboring disjoint alterations (GBM7-I1: L62R and R108K; GBM7-I2: A289V and C624S), suggesting that partial genetic information of a single tumor biopsy can be inconclusive for assessing EGFR targeted therapy (Supplementary Fig. 1b–c).

To understand the association between spatiotemporal architecture and genetic relevance, we calculated Nei's genetic distance among multi-sector samples from the same patient (Methods). Genetic diversity is greater in multiple tumors compared to locally-adjacent tissues (q=4.7e-5, Fig. 2a), in distant compared to local recurrence (q=1.4e-5, Fig. 2a), and in long-term recurrence compared to short-term recurrence (q=2.9e-3, Fig. 2a). A multinomial logistic regression was applied to classify multi-sector sample pairs based on genomic features (Supplementary Note). This analysis has highlighted that tumors from distant regions or longer period separation compose a distinct evolutionary scenario in GBM (Figure 2b). In colorectal tumors, a Big Bang model[20] interprets that cells from different biopsies of the same tumor share clonal and subclonal variants (Fig. 2c, left panel). Consistent with this model, samples taken from locally adjacent tumors share a large proportion of clonal and subclonal events (Supplementary Fig. 2a–b). In contrast, multiple tumors contain fewer shared (higher private) clonal mutations compared to local (q=1.86e-3, Supplementary Fig. 2a). We corroborated this finding by computing statistics on the space of evolutionary trees (evolutionary moduli spaces, Supplementary Fig. 3, Methods). Local tumors clustered near the tip of the space, indicating a higher shared mutation ratio compared to multiple tumors (p=1.27e-2). These results indicate that in contrast to local 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 panel). Unlike the Big Bang model[20], in the Multiverse model, tumor samples from different tumor masses share very few genomic alterations, indicating tumor clones are geographically segregated at an early stage of evolution, and each clone acquires distinct "private" alterations, leading to the construction of multiple "universes".

Next, we investigated the mutation profiles of GBMs with multifocal/multicentric (M-GBM) or solitary (S-GBM) lesion in 160 treatment naïve patients from both SMC and TCGA[17] cohorts (Supplementary Fig. 4 and 5a, Supplementary Table 3, Methods). Notably, non-synonymous mutations of *PIK3CA* were enriched in M-GBM (13/130 and 9/30 tumors in S- and M-GBMs respectively, p=7.905e-3, Fig. 2d). This conclusion remains the same in *IDH1* wild type cohort (Supplementary Fig. 5b). *PIK3CA* induces multipotency of mammary tumors, suggesting its associative role in tumor multiplicity[21]. Survival analysis indicates that both M-GBM and *PIK3CA* mutant patients have worse prognosis (p-values: 0.0151 and 0.039, respectively, Supplementary Fig. 5c–d).

To further characterize the heterogeneity of expression profiles, we curated single cell RNA-Seq of seven different samples from three patients. Overall, expression-based cell subtypes were not clearly determined by location or time (Fig. 3), consistent with a previous report[10]. To make sure this observation is not due to the limitations of this classification and to

capture the transcriptional similarity among different cells, we used topological data analysis (Methods), a recently developed technique to summarize and reduce the dimensionality of large data sets while retaining local high dimensional structure[22,23].

GBM9 (Figs. 3a–d) consisted of samples from two primary 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 the left initial tumor cells (Supplementary Figs. 6–8, Fig. 3b). Particularly, there are 61 somatic mutations shared between left and recurrent, while only 42 between right and recurrent. Single-cell transcriptome analysis showed *EGFR* expression predominantly in the right tumor mass, but not in the left and recurrent tumors (Fig. 3c). Interestingly, different single cells harbor different *EGFR* alterations, implying they were late events during tumor evolution[21]. *PIK3CA* mutations were detected from single cells in all three samples, consistent with the bulk WES that *PIK3CA* mutations are founder events (Supplementary Figs. 1a and 6a). Our analysis also revealed the presence of transcriptional heterogeneity within the individual samples. A subset of left-initial tumor cells was characterized by upregulation of mitotic genes not found in either right or recurrent sections (Fig. 3d).

Additionally, we profiled *IDH1* mutant tumor cells distinguished by 5-Aminolevulinic acid (5-ALA) uptake pattern (populations stained for tumor cellularity[24]) (GBM10, Fig. 3e–f, Supplementary Figs. 6–7). Previous glioma studies suggested that low pathologic grade is associated with low uptake of 5-ALA[24,25]; however, genomic determinants for 5-ALA uptake remain elusive. Using single cell transcriptome analysis, we found predominant enrichment of proneural cells on 5-ALA (-) sample, supporting previous observations that GBM cells may be evolved from proneural precursors (p<0.01, Fig. 3e,)[26]. We also found enrichment of several cell proliferation and migration markers in the 5-ALA (+) section, including *MET* and *CD44*[27–29]. Notably, 5-ALA (-) tumors, which are conceived as less aggressive, are actually fully-fledged tumors that harbor driver mutations and express tumor aggressive markers (Fig. 3e and 3f, Supplementary Fig. 6b, 7, 9a and 10). Finally, we studied main and 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, Supplementary Figs. 6–7, Supplementary Figs. 9–10)[28].

To investigate the influence of genetic heterogeneity on drug response, we isolated 28 PDCs from 11 patients 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; Fig. 4a). Consistently, both distant and longitudinal samples showed significantly broader drug responses compared to local samples (Fig. 4b). Importantly, we found that the PDCs from M-GBMs were more sensitive to PI3K/AKT/mTOR (PAM) pathway inhibitors than those from solitary tumors (p=1.872e-6, Fig. 4c and Supplementary Fig. 11). This indicates that PAM inhibitors could provide clinical benefit for M-GBMs. In addition, we

observed that PDCs from recurrent GBMs were more resistant to EGFR inhibitors, compared to initial (p=2.9e-4, Supplementary Figs. 12–13).

We hypothesized that clonal alterations found in all multi-sector samples (truncal alterations) represent better molecular targets. In agreement with this truncal target hypothesis (Supplementary video), multisector PDCs were more sensitive to drugs that target shared alterations compared to private alterations (p=0.0381, Methods, Fig. 4d; 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 in the left side tumor. However, PAM pathway inhibitors were ubiquitously effective, 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 no longer required for tumor maintenance, are not good candidates[31]. Although targeting subclonal mutations show limited effect, patients might still benefit from the elimination of a subclone that has a bystander effect on surrounding cells.

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

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 $5 \times 5 \times 5$ 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 GBM patients in 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 83 and 77 tumor exome sequencing data with matched normal DNA from SMC and TCGA[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 media with N2 and B27 supplements (0.5× each; Invitrogen) and 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-naïve GBMs (83 and 77 tumors from SMC and TCGA cohort, respectively) were reviewed. MR images of 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-GBM) from solitary ones (S-GBMs), we adapted annotations from the VASARI feature set for human glioma[36]. According to VASARI feature set, m-GBMs are defined as having at least one region of tumor, either enhancing or non-enhancing, which 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 FLAIR/T2 high signal intensity lesion resides outside of the T1-weighted contrast-enhancement lesion, this lesion is considered as separate tumor foci and counted as multifocal tumor in our study[39,40]. In contrast, tumors which present separate contrast-enhancement lesions within the FLAIR/T2 high-signal intensity background are considered as solitary ones.

Whole-exome sequencing

Agilent SureSelect kit was used for capturing the exonic DNA fragments. Illumina HiSeq2000 was used for sequencing and generated 2 × 101 bp paired-end reads.

Somatic mutation

The sequenced reads in 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 for refining the mutation calls from the above pipeline.

Copy number

Excavator was used to generate estimated copy number alterations in tumor specimens compared to its matching non-neoplastic part. For each gene we calculated the copy number $=2^{x+1}$, where x is the segmentation mean from Excavator and defined as the log2 fold-change of tumor divided by normal. 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 (CCF) and removed those having <20% purity. We

considered mutations as clonal if indicated clonal in Absolute and with a cancer cell fraction of 80% or having CCF of 100% and not marked as clonal or subclonal. The Absolute CCF estimates with regard to hypermutated samples appeared disproportionately subclonal in sample GBM18 inital and TCGA-14-1402 2nd recurrence and we reasoned the large mutational load may skew estimates. In hyper mutated samples treatment-associated mutation coupled with defects in mismatch repair are deemed largely responsible for a majority of observed mutations. Therefore, mutations having cancer cell fractions 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. Based on Figure 7b from their paper[42], 13 mutations have high cell fraction (>60%) shared by all samples. To relate the number of sectors sequenced to the number of mutations deemed to be clonal tumor-wide, we exhausted all possible sub-sampling 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 out of 36 sampling strategies contained no false discoveries in identifying clonal mutation. For each value of k, we calculated 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 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 distance of CCF for each patient's sample as follows. Let X = all CCF of sample 1 and Y = all CCF of sample 2:

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

The multiverse model of tumor evolution

We found increased Nei's genetic distance in Multi-focal/multi-centric compared with locally-adjacent biopsies. In addition, private clonal mutations appear frequently in multi-sectional and distant longitudinal samples, but are infrequent in local samples (Supplementary Figure 2a). This spurred a hypothesis that specific early event(s) can give rise to distinct mutational profiles in spatially separated tumors (Figure 2A). These differences in mutational load suggested distinct tumor profiles may arise in separate 'universes' of clones rather than from one large growth period followed by diversification.

For each somatic mutation we record the clonal status as determined by ABSOLUTE and if the mutation is shared or private or if the clonal status changes between biopsies. Mutations are then classified into 5 patterns between every available pair of a patient's samples. The mutational classes are labeled as the following: CC (clonal-clonal), CS (clonal-subclonal), SS (subclonal-subclonal), CX (clonal-absent) or SX (subclonal-absent). Order of the sample pair is not important: a mutation that is clonal in one sample and subclonal in the other is 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 multi-sectional/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 multi-sectional/distant-longitudinal, locally adjacent or local longitudinal sections in agreement with most of our MRI classifications. The simplex axes represent the predicted probabilities of outcomes for each observation. Curiously, the sample layout contained 3 longitudinal local outliers closest to the multi-sectional point of the simplex. The time interval between surgeries of the 3 pairs was 18 months or more. Moreover, their Nei distances were significantly different from all other sections (p-value = 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 (https://CRAN.R-project.org/package=nnet).

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. 352R and L cells were captured as a single isolate in C1 chip (17–25 μm) determined by microscopic examination as previous described. RNAs from pooled samples were also processed using the SMARTer kit with 10ng of starting materials. Libraries were generated using Nextera XT DNA Sample Prep Kit (Illumina) and sequenced on the HiSeq 2500 using the 100bp 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. TPM values were calculated from each single cell (as if they are different samples) using RSEM (version 1.2.25)[43] and expressed as $\log_2(1+\text{TPM})$.

Gene fusion detection

Chimerascan was applied to generate candidate list of gene fusions[44]. For bulk sequencing, only previously reported in-frame, high expressing fusions, such as *FGFR3-TACC3*[18], *MGMT* fusion[14], *EGFR-SEPT14*[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 will be reported.

Expression based subtypes determination

Gene expression was measured by RSEM and then log2 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) on the normalized expression profile. For each cell, we ranked all genes based on their expression values to create a .rnk file as the input of 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 transcriptome

We filtered out normal cells based on their expression profile. To that end, we considered expression signatures of normal oligodendrocytes, neurons, and astrocytes [45], microglia [46], endothelial cells [47], T-cells [48], and other immune cells [48,49], and used a Gaussian mixture model to classify individual cells according to their expression profile. 94/133, 82/85 and 90/137 cells, respectively for GBM9, GBM10, and GBM2, were classified as tumor cells. After normalization of gene expression level by dividing total number of reads in each cell to eliminate the potential bias caused by batch effect, we built topological representations of these single cell data using Mapper algorithm [23], as implemented by Ayasdi Inc. Open-source implementations of this algorithm are also available (http://danifold.net/mapper, http://github.com/MLWave/kepler-mapper). 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 representations of the data, where nodes represent sets of cells with similar global transcriptional profiles (as measured by the correlation of the expression levels of the 2,000 genes with highest variance across each patient). We identified individual genes that had an expression pattern localized in the network, and used those to determine the sub-clonal 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 the plating, PDCs were treated with drugs in a fourfold and seven-point serial dilution series from 20 μM to 4.88 nM using Janus Automated Workstation (PerkinElmer, Waltham, MA, USA). After 6 days of incubation at 37°C in a 5% CO2 humidified incubator, cell viability was analyzed using an adenosine triphosphate (ATP) monitoring system based on firefly luciferase (ATPLite™ 1step, PerkinElmer). Viable cells were estimated using EnVision Multilabel Reader (PerkinElmer). Dimethyl sulfoxide (DMSO) was also included as control in each plate. Controls were used for calculation of relative cell viability for each plate and normalization per-plate basis. DRC fitting was performed using GraphPad Prism 5 (GraphPad) and evaluated by measuring Area Under the Curve (AUC) of dose response curve. 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 dose–response curves (DRCs) with exclusion of non-convergent fits[50].

Moduli space analysis

To illustrate evolution histories of GBM patients, we applied moduli space analysis[51] in local and multiple group of patients. Multiregion pairs were compared to calculate number of shared and private mutations. In this analysis, clonal mutations were separated based on their allele frequencies. Sector pairs were put in left sphere based on the number of shared and private mutations with high allele frequency mutations (>20%), while the same number of pairs were put in right sphere based on mutations with low allele frequency mutations (<20%). The same analysis was then also performed based on inferred clonally.

Immunohistochemistry

Tissue specimens were fixed by formalin and embedded in paraffin. Paraffin-embedded sections were treated with 0.3% Hydorgen peroxide to block endogenous peroxidase acitivty and antigens were retrieved by heating sections in 10 mM sodium citrate (pH 6.0) for 95 degree Celsius 30 min. Sections were incubated with primary antibodies overnight at 4 degree Celsius, biotinylated secondary antibodies for 1 hr in RT and avidin-biotin complex for 1 hr in RT.

Western blot

GBM PDCs were washed with cold PBS, 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 added (Thermo Scientific). Insoluble materials were removed by centrifugation at 12000 rpm for 15 min, 4°C. The proteins were separated by SDS-PAGE. Immunoblotting was performed using antibodies against indicated proteins.

Limiting dilution assay

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 one to two 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 reverse transcription polymerase chain reaction (RT-PCR) assays. Total RNA was extracted from the 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 a random primer using the SuperScriptIII First-Strand System(Life Technologies), and 20 μl of synthesized cDNA was diluted 5 fold with DW. For RT-PCR, EzWay Taq PCR MasterMix (Komabiotech, KOREA) and 5 μl of synthesized cDNA template was used. Thermal cycling was carried out under the following conditions: 1 min at 95°C followed by 30 cycles of 30 sec at 95°C, 30 sec at 56°C, 30 sec at 72°C. The primer pairs used in this experiment were designed to make the amplification product including the breakpoints of the fusion genes. PCR products were

analyzed by agarose gel electrophoresis. The primers were summarized in Supplementary Table 6.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 following sources: 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).

References for main text

1. Hamburg MA, Collins FS. The path to personalized medicine. N Engl J Med. 2010; 363:301–4. [PubMed: 20551152]
2. Brennan CW, et al. The somatic genomic landscape of glioblastoma. Cell. 2013; 155:462–77. [PubMed: 24120142]
3. Ceccarelli M, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell. 2016; 164:550–63. [PubMed: 26824661]
4. Cloughesy TF, Cavenee WK, Mischel PS. Glioblastoma: from molecular pathology to targeted treatment. Annu Rev Pathol. 2014; 9:1–25. [PubMed: 23937436]
5. Frattini V, et al. The integrated landscape of driver genomic alterations in glioblastoma. Nat Genet. 2013; 45:1141–9. [PubMed: 23917401]
6. Snuderl M, et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell. 2011; 20:810–7. [PubMed: 22137795]
7. Szerlip NJ, 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 U S A. 2012; 109:3041–6. [PubMed: 22323597]
8. Kim H, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. 2015; 25:316–27. [PubMed: 25650244]
9. Sottoriva A, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A. 2013; 110:4009–14. [PubMed: 23412337]
10. Patel AP, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014; 344:1396–1401. [PubMed: 24925914]
11. Kumar A, et al. Deep sequencing of multiple regions of glial tumors reveals spatial heterogeneity for mutations in clinically relevant genes. Genome Biol. 2014; 15:530. [PubMed: 25608559]
12. Johnson BE, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science. 2014; 343:189–93. [PubMed: 24336570]
13. Yates LR, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat Med. 2015; 21:751–9. [PubMed: 26099045]
14. Wang J, et al. Clonal evolution of glioblastoma under therapy. Nat Genet. 2016
15. Ohka F, Natsume A, Wakabayashi T. Current trends in targeted therapies for glioblastoma multiforme. Neurol Res Int. 2012; 2012:878425. [PubMed: 22530127]
16. Kim J, et al. Spatiotemporal Evolution of the Primary Glioblastoma Genome. Cancer Cell. 2015; 28:318–328. [PubMed: 26373279]
17. Liu Q, et al. Genetic, epigenetic, and molecular landscapes of multifocal and multicentric glioblastoma. Acta Neuropathol. 2015; 130:587–97. [PubMed: 26323991]

18. Singh D, et al. Transforming fusions of FGFR and TACC genes in human glioblastoma. Science. 2012; 337:1231–5. [PubMed: 22837387]
19. Hartmann C, Bartels G, Gehlhaar C, Holtkamp N, von Deimling A. PIK3CA mutations in glioblastoma multiforme. Acta Neuropathol. 2005; 109:639–42. [PubMed: 15924253]
20. Sottoriva A, et al. A Big Bang model of human colorectal tumor growth. Nat Genet. 2015; 47:209–16. [PubMed: 25665006]
21. Koren S, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature. 2015; 525:114–8. [PubMed: 26266975]
22. Carlsson G. Topology and data. Bulletin of the American Mathematical Society. 2009; 46:255–308.
23. Singh, G., Mémoli, F., Carlsson, GE. SPBG. Citeseer: 2007. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition; p. 91-100.
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. 2014; 14:146.
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. 2006; 7:392–401. [PubMed: 16648043]
26. Ozawa T, et al. Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell. 2014; 26:288–300. [PubMed: 25117714]
27. Merzak A, Koocheckpour S, Pilkington GJ. CD44 mediates human glioma cell adhesion and invasion in vitro. Cancer Res. 1994; 54:3988–92. [PubMed: 7518347]
28. Yoshida T, Matsuda Y, Naito Z, Ishiwata T. CD44 in human glioma correlates with histopathological grade and cell migration. Pathol Int. 2012; 62:463–70. [PubMed: 22726066]
29. Joo KM, et al. MET signaling regulates glioblastoma stem cells. Cancer Res. 2012; 72:3828–38. [PubMed: 22617325]
30. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010; 141:1117–34. [PubMed: 20602996]
31. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996; 87:159–170. [PubMed: 8861899]
32. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010; 17:98–110. [PubMed: 20129251]
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. 2006; 9:391–403. [PubMed: 16697959]
34. Clark K, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013; 26:1045–57. [PubMed: 23884657]
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. 2015; 17:1525–37. [PubMed: 26203066]
36. Rios Velazquez E, et al. Fully automatic GBM segmentation in the TCGA-GBM dataset: Prognosis and correlation with VASARI features. Sci Rep. 2015; 5:16822. [PubMed: 26576732]
37. Giannopoulos S, Kyritsis AP. Diagnosis and management of multifocal gliomas. Oncology. 2010; 79:306–12. [PubMed: 21412017]
38. Kyritsis AP, Levin VA, Yung WK, Leeds NE. Imaging patterns of multifocal gliomas. Eur J Radiol. 1993; 16:163–70. [PubMed: 8389708]
39. Zinn PO, et al. Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. PLoS One. 2011; 6:e25451. [PubMed: 21998659]
40. Li YM, 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. 2016; 124:977–88. [PubMed: 26495941]
41. Carter SL, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol. 2012; 30:413–21. [PubMed: 22544022]

42. Suzuki H, et al. Mutational landscape and clonal architecture in grade II and III gliomas. Nature Genetics. 2015; 47:458–68. [PubMed: 25848751]

43. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. [PubMed: 21816040]

44. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics. 2011; 27:2903–2904. [PubMed: 21840877]

45. Cahoy JD, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci. 2008; 28:264–78. [PubMed: 18171944]

46. Butovsky O, et al. Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. Nat Neurosci. 2014; 17:131–43. [PubMed: 24316888]

47. Bhasin M, et al. Bioinformatic identification and characterization of human endothelial cell-restricted genes. BMC Genomics. 2010; 11:342. [PubMed: 20509943]

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. 2005; 175:7837–47. [PubMed: 16339519]

49. Abbas AR, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes Immun. 2005; 6:319–31. [PubMed: 15789058]

50. Huang S, Pang L. Comparing statistical methods for quantifying drug sensitivity based on in vitro dose-response assays. Assay Drug Dev Technol. 2012; 10:88–96. [PubMed: 22066911]

51. Ji, L., Yau, S-T. Transformation groups and moduli spaces of curves. International Press; Somerville, Massachusetts: 2011. p. iip. iiip. 299

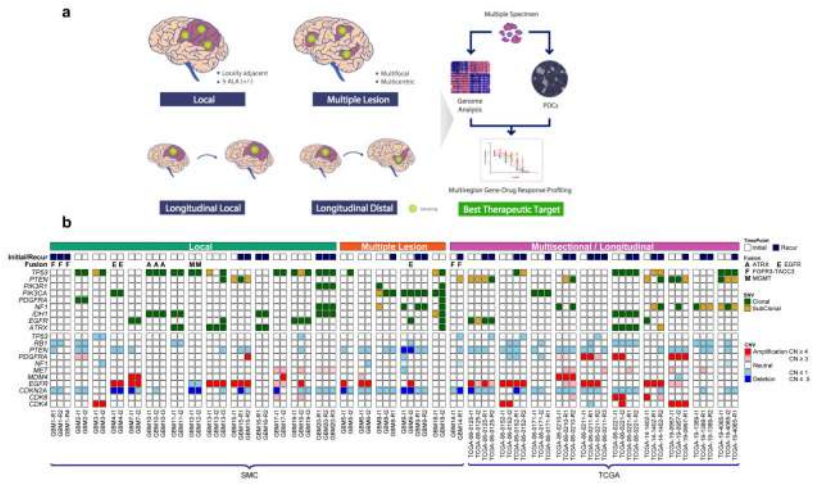

Figure 1. Mutational landscape of multi-region malignant glioma samples

(a) A schematic representation of glioma genomic heterogeneity and differential drug response analysis. Human glioma specimens were acquired based on their spatial order, or longitudinal pairs and subjected for genomic analysis for identification of tumor-initiating (truncal) events. (**b)** Somatic mutations including Single Nucleotide Variants (SNVs) and small Insertions/Deletions, copy number alterations, and gene fusions of 83 glioma multi-region or multisector-longitudinal specimens from 30 patients are demonstrated. 34 locally adjacent tumor fragments were from 14 patients, 13 multifocal/multicentric (referred as multiple) tissues from 5 patients, and a longitudinal pair GBM14 with leptomeningeal seeding were collected from Samsung Medical Center (SMC). We also curated 34 multisector-longitudinal tumor exomes and/or RNA sequencing from 10 patients in TCGA cohort[8]. All somatic mutations called by SAVI with allele frequency >5% were demonstrated. For each gene we calculated the copy number (CN) based on Excavator. Clonal alterations were determined using ABSOLUTE with cancer cell fraction >80%. (Methods).

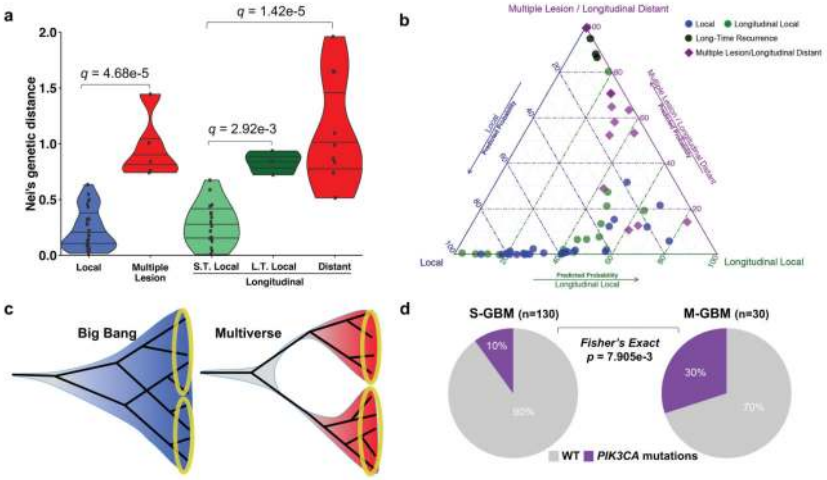

Figure 2. Comparison of genetic heterogeneity across glioma multisector / longitudinal samples

Patient samples were classified into five groups: Local, Multiple Lesion, S.T. (Short-Term) Longitudinal Local, L.T. (Long-Term) Longitudinal Local and Longitudinal Distant for comparative analyses. (**a)** Nei's genetic distance of the indicated groups are shown. Q-values were calculated by Wilcoxon Rank-Sum Test and corrected for false discovery rate using Benjamini-Hochberg method. S.T. and L.T. Local indicates short-time (<18 months surgical interval) and long-time recurrent tumors (≥18 months), respectively. **(b)** Illustration of leave-one-out results from multinomial logistic regression. Each point indicates one pair of samples, and their coordinates are the probabilities to be local, multiple lesion/longitudinal distant, or longitudinal local. Long-Time recurrent samples were classified together with multiple lesion/longitudinal distant samples, indicating they might follow the same evolutionary model. **(c)** Tumor evolution behind Big Bang and Multiverse models. The Big Bang model is indicated as a mixture of tumor cells that share many clonal and subclonal alterations. The Multiverse model is indicated with a greater proportion of private events at a clonal level. **(d)** Pie charts demonstrate the frequencies of *PIK3CA* mutations in multifocal/ multicentric (M-) GBMs (30%, 9/30) and solitary (S-) GBMs (10%, 13/130). The *p*-value was calculated using Fisher's exact test.

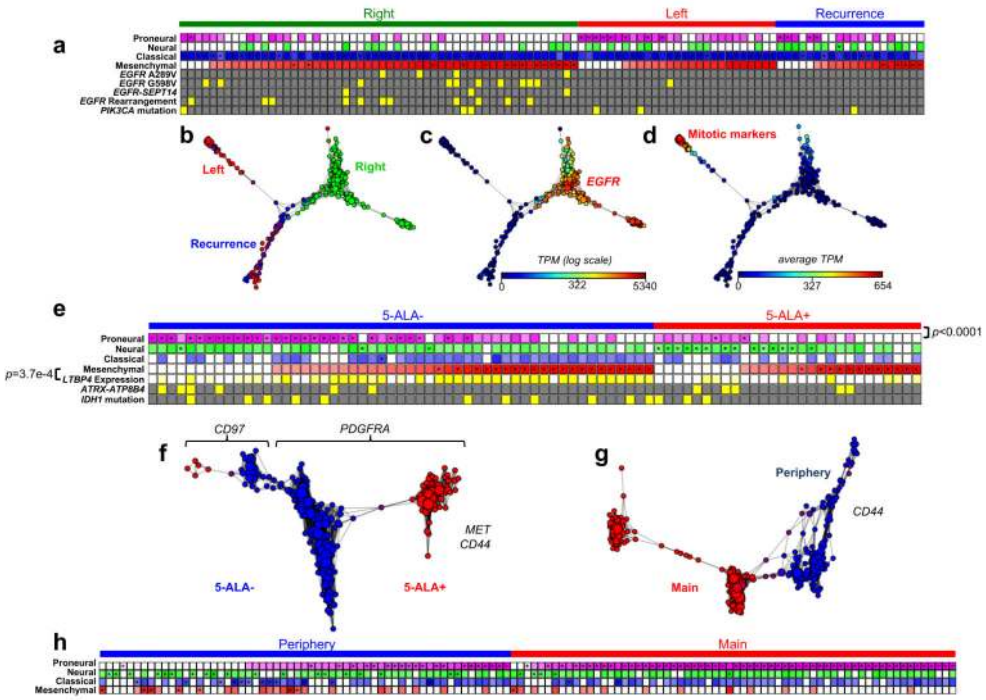

Fig. 3. Single cell transcriptome from multi-region samples

(a) Expression profile of individual tumor cells from three samples of GBM9 (left initial, right initial, and relapse), according to expression subtypes[32]. For each cell, the subtype with the highest expression is marked with an asterisk. Several *EGFR* genomic alterations can 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 profile. A cell can appear in several nodes, and two nodes are connected by an edge if they have at least one cell in common. Topological representation of GBM9 labeled by expression of *EGFR* **(c)** and mitotic genes **(d). (e)** Expression profile of individual tumor cells from GBM10 (two samples: 5-ALA (+) and 5-ALA (-)). The *p*-value between proneural and 5-ALA was obtained using Fisher's exact test. The *p*-value between mesenchymal and *LTBP4* expression was calculated based on Spearman's correlation. *ATRX* fusion was validated by RT-PCR assays (Supplementary Fig. 7b–c). Topological representation of expression data of individual tumor cells from patients GBM10 **(f)** and GBM2 **(g)**, labeled by the sample of origin. **(h)** Expression profile of individual tumor cells from patient GBM2 according to the GBM expression subtypes.

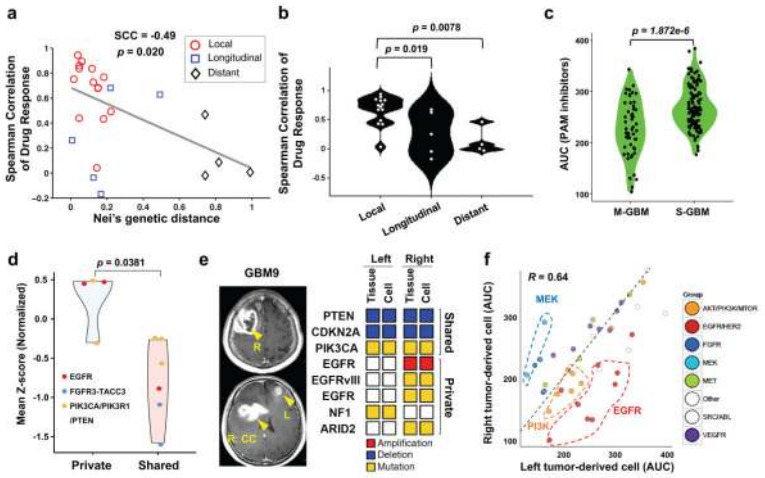

Fig 4. Chemical screening of multi-region patient-derived cells (PDCs)

(a) PDCs were treated with 40 chemical agents, targeting oncogenic signaling pathways in diluted series from 20 μM to 4.88 nM. X-axis indicates Nei's genetic distances between fragments from the same patient, while y-axis indicates Spearman's correlation coefficient (SCC) of corresponding fragments based on drug sensitivities measured by Area Under the Curve (AUC). **(b)** A violin plot for SCC of drug responses of the groups described in Fig. 4a. **(c)** Mean values of AUCs for six PI3K/AKT/mTOR (PAM) inhibitors (BEZ235, BKM120, BYL719, AZD5363, AZD2014 and Everolimus) of PDCs isolated from M- (n=9) or S-GBMs (n=22) were plotted. **(d)** The normalized Z-score in each PDC was plotted when the corresponding tissues harbored associated genetic alterations, designated as "shared" or "private". The private group was determined when the drug response-associated genetic alteration (i.e. *EGFR* mutations-EGFR inhibitors; *PTEN* mutations-PI3K/AKT/mTOR pathway inhibitors) was private, and vice versa for the shared group. **(e)** Preoperative T1-weighted contrast-enhanced MR image and key genomic alterations found in the corresponding tumors and its derivative cells from a multicentric patient (GBM9). Right-side 'R' tumors encompassed the right frontal lobes and corpus callosum (CC). 'L' indicates the left frontal tumor. Preoperative MRI showed a multifocal infiltrative lesion in both the frontal lobe and CC. **(f)** Scatterplot of AUC for 40 cancer-targeting compounds on GBM9 PDCs derived from the left and right side tumor. The R was obtained as Pearson's correlation coefficient. All *p*-values in this figure were obtained using Wilcoxson Rank-Sum test.