Longitudinal assessment of tumor development using cancer avatars derived from genetically engineered pluripotent stem cells

Tomoyuki Koga 1,2,9, Isaac A. Chaim 3,5,9, Jorge A. Benitez 1, Sebastian Markmiller 3, Alison D. Parisian 1,4, Robert F. Hevner 4, Kristen M. Turner 1, Florian M. Hessenauer 1, Matteo D’Antonio 5, Nam-phuong D. Nguyen 6, Shahram Saberi 7, Jianhui Ma 1, Shunichiro Miki 1, Antonia D. Boyer 1, John Ravits 7, Kelly A. Frazer 5,8, Vineet Bafna 6, Clark C. Chen 2, Paul S. Mischel 1,4, Gene W. Yeo 3,5* & Frank B. Furnari 1,4*

Many cellular models aimed at elucidating cancer biology do not recapitulate pathobiology including tumor heterogeneity, an inherent feature of cancer that underlies treatment resistance. Here we introduce a cancer modeling paradigm using genetically engineered human pluripotent stem cells (hiPSCs) that captures authentic cancer pathobiology. Orthotopic engraftment of the neural progenitor cells derived from hiPSCs that have been genome-edited to contain tumor-associated genetic driver mutations revealed by The Cancer Genome Atlas project for glioblastoma (GBM) results in formation of high-grade gliomas. Similar to patient-derived GBM, these models harbor inter-tumor heterogeneity resembling different GBM molecular subtypes, intra-tumor heterogeneity, and extrachromosomal DNA amplification. Re-engraftment of these primary tumor neurospheres generates secondary tumors with features characteristic of patient samples and present mutation-dependent patterns of tumor evolution. These cancer avatar models provide a platform for comprehensive longitudinal assessment of human tumor development as governed by molecular subtype mutations and lineage-restricted differentiation.
Effective modeling of cancer has been a conceptual cornerstone in the field of oncology for studying pathobiology and identifying therapeutic targets. In the case of glioblastoma (GBM), the most common primary malignant tumor of the central nervous system, mouse models of GBM-like tumors generated through the genetic disruption of different combinations of core tumor suppressors and/or by introduction of oncogenes such as Src, K-ras, H-ras, PDGFB, and EGFRvIII are available to investigate the biology of these aggressive tumors or to test possible treatments in preclinical settings. These mouse models are suitable to investigate pathobiology of genetically defined gliomas and useful for drug testing, but typically lack the intra-tumor heterogeneity that is observed in human gliomas. In addition, while human astrocytes engineered with combinations of human TERT and H-Ras expression and inhibition of the TP53 pathway either by SV40 T/t-Ag or by HPV E6 and E7 generate gliomas with high-grade histology and sensitivity to pathway-specific therapies, however, they do not allow for experimental standardization or afford analysis of the effects of molecular subtype mutations on tumor evolution.

The progress in human stem-cell technologies and genome editing using site-specific nucleases such as ZFN, TALEN, and CRISPR/Cas9 has broadened the field of human disease modeling. Such engineering has also been efficiently applied to neural stem cells providing opportunities for functional genetic analysis. This combination of human stem cell and genome editing promises great potential when applied to cancer models. The first such model generated utilized colon organoids derived from human intestinal crypt stem cells engineered with four or five mutations common in colorectal cancers. These organoid models accurately predict drug responses and their utility is anticipated for application of personalized therapies. Later, a brain tumor model deleted for PTEN by TALEN-mediated homologous recombination led to the reprogramming of human neural stem cells toward a cancer stem cell-like phenotype. However, it remains unknown if these cancer models generated through genome editing harbor authentic pathological features of cancers, including tumor heterogeneity and clonal evolution.

Here, we establish a robust platform in an isogenic background, which uses CRISPR/Cas9 genome editing technology and serial in vivo engraftments enabling longitudinal assessment of human high-grade glioma (HGG) models containing combinations of genetic alterations observed in proneural and mesenchymal GBM molecular subtypes. We further present how closely these models recapitulate pathobiology of the disease and discuss their utility as an avatar platform for future studies on tumor biology and evolution.

**Results**

**Neural progenitors with GBM mutations form HGG-like tumors.** We first introduced two different combinations of driver mutations into human induced pluripotent stem cells (iPSCs) by CRISPR/Cas9 genome editing. (Fig. 1a, b). One combination of deletions targeted tumor suppressor genes PTEN and NF1, which are commonly altered together in the mesenchymal subtype of GBM. A second combination of deletions targeted TP53 and exons 8 and 9 of PDGFRα (PDGFRαΔ8-9). This creates a constitutively active truncating PDGFRα mutation observed in 40% of PDGFRα amplified GBM, resulting in a genotype commonly found in the proneural subtype of isocitrate dehydrogenase-wildtype GBM. The genetic modifications in single clones were confirmed by genotyping PCR (Fig. 1c) and RT-qPCR (Fig. 1d). Edited iPSC clones with desired mutations were differentiated into neural progenitor cells (NPCs), using a small molecule protocol and differentiation status was confirmed by downregulation of pluripotency markers, Nanog and Oct4, and corresponding upregulation of NPC markers, Pax6, Nestin, and Sox1 (Fig. 1e). These edited NPCs were expanded on matrigel-coated plates in NPC maintenance media and were utilized in further experiments.

We next evaluated if these genetically modified NPCs were capable of forming orthotopic tumors in immunocompromised mice (Fig. 1a). When edited NPCs were engrafted in the brains of four Nod scid mice, PTENΔ8-9;NF1−/− NPCs and TP53Δ8-9;PDGFRαΔ8-9 NPCs each formed brain tumors with median survival of 141, and 119.5 days, respectively (Fig. 1f). Pathological assessment of PTENΔ8-9;NF1−/− tumors revealed regions of hypercellularity with occasional mitoses (Fig. 2a), and in one out of four tumors, there were biphasic dense glial and loose mesenchymal/sarcoma morphologies, typical of gliosarcoma (Fig. 2b). In addition, regions of necrosis (Fig. 2c), vascular endothelial proliferation (Fig. 2d), subarachnoid spread (Fig. 2e), perineuronal satellitosis, and subpial accumulation of tumor cells (Fig. 2f), were also apparent. The tumors were consistently positive for GFAP (3+ in 6/6 high power fields) (Fig. 2g, Supplementary Fig. 2b), and also highly positive for Ki-67 staining (Fig. 2h, Supplementary Fig. 2b), and PDGFRαΔ8-9 tumors presented nodular growth of a primitive neuronal component (dark purple) intermingled with glial components (Fig. 2i), rosettes with neurupil-like texture (Fig. 2j), a serpiginous zone of pseudopalisading necrosis (Fig. 2k), and intraventricular growth (Fig. 2l). These tumors were positive for GFAP (3+ in 6/6) (Fig. 2m, Supplementary Fig. 2a) and Olig2 (3+ in 6/6 high power fields, 2+ in 2/6 high power fields) (Fig. 2n, Supplementary Fig. 2a). In terms of WHO grade, three and one out of four tumors were scored as grade 4 and grade 3, respectively for PTENΔ8-9;NF1−/− tumors, and four out of four tumors were scored as grade 4 for TP53Δ8-9;PDGFRαΔ8-9 tumors (Supplementary Fig. 3). In contrast, PTENΔ8-9 and TP53Δ8-9 singly edited NPCs did not form tumors in the brain over the same time span (Fig. 1f, Supplementary Fig. 4a, b), while unedited iPSCs formed teratoma-like tumors (Supplementary Fig. 5). Lack of teratomas after NPC injection suggests a high efficiency of differentiation to NPCs. These results illustrate that using this modeling paradigm, small numbers of known driver mutations found in GBM are sufficient for phenotypic recapitulation of human HGG tumors.

**iHGG cells can be cultured and form secondary tumors.** One of the benefits of using PDX models in cancer research is that they can be cultured in vitro and be re-engrafted in animals, thus enabling both in vitro and in vivo analyses. We evaluated if our induced HGG (iHGG) models could be used in a similar manner. Dissociated tumors obtained from the mouse brains were sorted into individual tumor spheres that were capable of forming orthotopic tumors in immunocompromised mice, when compared to pre-engraftment NPCs (Fig. 3b), again highlighting gain of cancerous phenotypes of iHGG cells compared to original input cells.
We then evaluated if these iHGG-derived sphere cells maintained tumorigenic capacity by secondary orthotopic engraftment (Fig. 3c). When injected in the brains of Nod scid mice, PTEN−/−;NFI−/− and TP53−/−;PDGFRαΔ8–9 iHGG-derived sphere cells formed tumors with a shortened latency period of median 76.5 days and 34.5 days, respectively (p = 0.0005, log-rank test) (Supplementary Fig. 7). We also tested if these models can be used for in vivo drug treatment experiments comparable to those applied to PDX lines by treating orthotopically engrafted animals with temozolomide (TMZ), a DNA-alkylating chemotherapeutic agent used for standard care treatment of GBM patients. TP53−/−;PDGFRαΔ8–9 iHGGs proved to be more sensitive to TMZ compared to PTEN−/−;NFI−/−.
iHGGs (Fig. 3d). \( PTEN^{+/+},NF1^{+/+} \) iHGGs were found to express higher levels of \( O^6 \)-methylguanine DNA methyl transferase (MGMT) (Fig. 3e), which is associated with resistance to TMZ in GBM patients\(^{26} \), compared to \( TP53^{−/−},PDGFRA^{Δ8−9} \) iHGGs. An alternative explanation of this differential sensitivity, through MGMT-independent mechanisms in the context of \( TP53 \) alteration\(^{27,28} \) cannot be eliminated.

iHGGs recapitulate molecular and genetic hallmarks of GBM. We further investigated if these iHGGs showed inter- and intra-tumor heterogeneity, which is another hallmark of GBM and cancer in general\(^{29} \). This important feature of cancer has not been well studied in previous models to date. To investigate the robustness of our iHGG models we performed in triplicate, single-cell RNA sequencing (scRNA-seq) using primary iHGG...
**Fig. 2** Histology of iHGGs. H&E staining of PTEN−/−;NFI−/− iHGGs showing a region of hypercellularity infiltrated by irregular, elongated to angulated tumor cells with occasional mitoses (a), scale bars, 50 μm (left) and 10 μm (right), biphasic dense (glial) and loose (mesenchymal/sarcoma) morphologies, typical of gliosarcoma (b), necrosis (central pink zone) with peripheral “pseudopalisading” of cells around the necrotic center (c), scale bars, 200 μm (b, c), vascular endothelial proliferation (d), scale bar, 100 μm, rupture through the pial surface, and consequently subarachnoid spread (upper right) (e), scale bar, 200 μm, and “secondary structures” typical of glioma, including perineuronal satellitosis and subpial accumulation of tumor cells (f), scale bar, 50 μm. GFAP (g), Olig2 (h), Ki-67 (i) staining of PTEN−/−;NFI−/− iHGGs, scale bars, 100 μm (g-i). H&E staining of TP53−/−;PDGFRΑΔ8−9 iHGGs showing nodular growth of a primitive neuronal component (dark purple) intermingled with glial component (pink) (j), scale bar, 200 μm, rosettes with neuropil-like texture in a primitive neuronal component (k), scale bar, 50 μm, a serpigenous zone of pseudopalisading necrosis (l), scale bar, 200 μm, and a tumor rupture through ependyma illustrating intraventricular growth (m), scale bar, 500 μm. GFAP (n), Olig2 (o), Ki-67 (p) staining of TP53−/−;PDGFRΑΔ8−9 iHGGs, scale bars, 100 μm (n-p).

**Fig. 3** Cells from iHGG models can be cultured in vitro and re-engrafted to form secondary tumors with different drug response. a) iHGG spheres obtained by maintaining iHGG tumor cells in neurosphere culture conditions, scale bars, 2 mm. b) Extreme limiting dilution analysis of input NPCs and tumor-derived iHGG sphere cells. c) H&E staining of secondary tumors generated from re-engraftment of primary iHGG spheres, scale bars, 5 mm, 250μm, 5mm, 250μm, (left to right). d) In vivo survival assays of mice orthotopically engrafted with primary iHGG spheres upon treatment either with vehicle or temozolomide. Data are representative of six replicates, n = 6 animals for each treatment arm for each model. Data were analyzed by the log-rank test. e) MGMT expression levels in iHGG cells analyzed by semi-quantitative RT-qPCR. Data are representative of three replicates, n = 3. Data are represented as mean ± SD, analyzed by unpaired t-test. Source data are provided as a Source Data file.

spheres, secondary iHGG tumor cells obtained from orthotopic injection of the primary spheres, as well as secondary spheres derived by in vitro culture of the secondary tumor cells for both genotypes, for a total of 14 samples (Fig. 4a).

Visual analysis of all the samples by Uniform Manifold Approximation and Projection (UMAP) reveals clear structural stratification between primary and secondary spheres of the same genotype as well as between spheres and tumors (Fig. 4b). However, the greatest variation appears between the two iHGG models of different genotypes. This inter-tumor heterogeneity between iHGG models was not apparent in pre-engraftment NPCs with different gene edits (Supplementary Fig. 8). In fact, the
strongest driver of transcriptomic differences is genotype, as shown by the clear split between them, regardless of their origin (spheres or tumors, primary or secondary) for the strongest principal component (PC), pc1 (Fig. 4c). These findings continue to support the notion that a small number of driver mutations are sufficient for the development of such pathognomonic inter-tumor heterogeneity that arises through the process of transformation.

Given that our models were engineered to recapitulate different clinical GBM molecular subtypes, specifically proneural and...
mesenchymal, we sought to determine if our samples manifested transcriptomic GBM signatures as established previously\(^8,18\). TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG show upregulation of genes characteristic of the proneural subtype, while the PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG show a mesenchymal subtype signature for both spheres and tumors (Fig. 4d, and Supplementary Fig. 9). Subtype scores involving all expressed genes under each subtype show similar trends, with higher proneural scores for TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG and higher mesenchymal scores for PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG (Supplementary Fig. 10). Interestingly, TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG samples also show increased classical subtype scores. Importantly, when examined at single-cell resolution, each sample shows intra-tumor heterogeneity with different populations of cells presenting signatures of different subtypes (Supplementary Fig. 10), as is characteristic of GBM patient samples\(^29\). In addition, all samples are comprised of populations of cycling and noncycling cells (Supplementary Fig. 11a, b), which is also characteristic of patient samples and is in juxtaposition with other in vitro GBM models, where almost 100% of cells are cycling\(^29\). Finally, in agreement with previous literature on patient samples, cells with high proneural scores also score highly on stemness as is the case for our TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG samples (Supplementary Fig. 11c–e)\(^29\). In conclusion, our results highlight the robustness of our iHGG models, for both spheres and tumors, in recapitulating hallmarks of patient GBM samples as is the case for cellular inter- and intratumor heterogeneity, subtype signatures and cycling and stemness scores.

**Distinct iHGGs present different patterns of tumor evolution.** It is also apparent from the UMAP plots that the transcriptomic signature of primary spheres evolves as they are passaged through mice, excised, and cultured in vitro. In fact, by analyzing separately the transition of each genotype model from primary to secondary spheres, we gain insights into the biology of the tumors as well as the role that in vivo passaging plays. We performed unsupervised Louvain clustering of PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG primary and secondary spheres (Fig. 5a) and, in parallel, TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG primary and secondary spheres (Fig. 5b) and found 8 and 15 distinct clusters, respectively. In both cases primary spheres are represented, almost exclusively, by unique clusters not found in any secondary spheres. Remarkably, all three PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG secondary spheres are found in very similar proportions in the remaining clusters. In stark contrast, each of the TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG secondary spheres show unique cluster makeups.

Moreover, when the differentially expressed genes of each cluster are subjected to gene ontology (GO) analysis, different patterns emerge for each iHGG model. PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG spheres are subdivided in two broad categories, namely cell cycle or cell motility and extra- and intra-cellular fiber reorganization, with several cluster sharing a variety of GO terms (Fig. 5c). In fact, both categories appear to be exacerbated following tumor formation in mice as shown by the increase in the number of GO terms associated with each secondary sphere cluster. This increase in cell motility terms supports our observation regarding more prominent diffuse invasion of PTEN\(^{-/-}\);NF1\(^{-/-}\) iHGG compared with TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) iHGG (Supplementary Fig. 12). In

---

**Fig. 5 Genetically distinct iHGG models present different patterns of longitudinal evolution.** Primary and secondary spheres of PTEN\(^{-/-}\);NF1\(^{-/-}\) (a) and TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) (b). UMAP plot color-coded by samples (top left) or by Louvain clustering (top right). Sample distribution found in each Louvain cluster, color-coded by sample identity (bottom left) and Louvain cluster distribution per sample, color-coded by cluster identity (bottom right). Clustered heatmaps of enriched GO terms extracted from differentially expressed genes of each Louvain cluster in PTEN\(^{-/-}\);NF1\(^{-/-}\) (c) and TP53\(^{-/-}\);PDGFRA\(^{Δ-9}\) (d). Color scale represents statistical significance. Gray color indicates a lack of significance.
NF1 labeling of chromosomes and ecDNA in a metaphase spread of

spreads and digital karyotyping, PTEN
tivities or traces of ecDNA. Based on DAPI staining of metaphase
gated if the original input NPCs possessed karyotype abnormal-

secondary spheres,
both iHGG models show transcriptional drift from primary to
in nature (Supplementary Figs. 13 and 14). Overall, even though

across most clusters, whereas proneural scores are heterogeneous
PTEN

signature of
motility associated terms (Fig. 5d). Likewise, the mesenchymal

glial cell differentiation, response to hypoxia, embryonic morpho-
genesis and, similarly to PTEN−/−;NF1−/− iHGG spheres, cell
motility associated terms (Fig. 5d). Likewise, the mesenchymal
signature of PTEN−/−;NF1−/− iHGG spheres is homogenous
across most clusters, whereas proneural scores are heterogeneous
in nature (Supplementary Figs. 13 and 14). Overall, even though
both iHGG models show transcriptional drift from primary to
tertiary spheres, TP53−/−;PDGFRAΔ8−9 iHGGs appear to show
a less unidirectional path with increased heterogeneity.

We previously reported that extrachromosomal DNA (ecDNA)
is prevalent in many cancer types, especially in GBM, and that
eDNA is associated with resistance to drug treatment and rapid
evolution of tumor heterogeneity. To determine if our iGBM
models recapitulated the generation of ecDNA, we first investi-
gated if the original input NPCs possessed karyotype abnormal-
ities or traces of ecDNA. Based on DAPI staining of metaphase
spreads and digital karyotyping, PTEN−/−;NF1−/− iHGG cells
were karyotypically normal (Fig. 6a). In sharp contrast, metaphase
spreads of cells obtained from TP53−/−;PDGFRAΔ8−9 iHGGs
showed small DAPI-stained dots adjacent to chromosomes,
suggestive of ecDNA (Fig. 6b), consistent with our previous
findings in GBM tumor samples. Furthermore, double minute-
like structures became more apparent in the secondary tumors
obtained by re-engraftment of the primary spheres (Fig. 6c), and
were replication competent as indicated by incorporation of EdU
(Fig. 6d). The TP53−/−;PDGFRAΔ8−9 iHGGs also presented
striking numerical and structural chromosome alterations (Fig. 6e).
This supports a clonally unstable nature of the TP53−/−;
PDGFRAΔ8−9 model, where genomic instability or ecDNA could
be driving dynamic accelerated clonal evolution.

iHGGs confirm features characteristic of patient samples. We
also applied unsupervised Louvain clustering of PTEN−/−;
NF1−/− iHGG secondary tumors (Fig. 7a) and, in parallel,
TP53−/−;PDGFRAΔ8−9 iHGG secondary tumors (Fig. 7b) to
further compare inter- and intra-tumor variability of our iHGG
tumor models and found 7 and 8 distinct clusters, respectively.
Each one of the clusters is represented by a unique set of

![Fig. 6 TP53−/−;PDGFRAΔ8−9 iHGG shows prominent karyotype abnormalities accompanied by extrachromosomal DNA. a DAPI staining of PTEN−/−;NF1−/− primary iHGG cells, scale bar, 10 μm. b DAPI staining of TP53−/−;PDGFRAΔ8−9 primary iHGG cells. Red arrows indicate ecDNA, scale bars, 10 μm (left), 2 μm (right). c DAPI staining of TP53−/−;PDGFRAΔ8−9 secondary iHGG cells. Red arrows indicate ecDNA, scale bars, 10 μm (left), 2 μm (right). d EdU labeling of chromosomes and ecDNA in a metaphase spread of TP53−/−;PDGFRAΔ8−9 secondary iGBM, scale bar, 5 μm. e Spectral karyotyping analysis of TP53−/−;PDGFRAΔ8−9 iHGG cells.](image-url)
**Fig. 7 iHGG tumors confirm features characteristic of patient tumor samples.** *PTEN*⁻/⁻;*NF1*⁻/⁻ (a) and *TP53*⁻/⁻;*PDGFRAΔB-9 (b) secondary tumors. UMAP plot color-coded by samples (top left) or by Louvain clustering (top right). Sample distribution found in each Louvain cluster, color-coded by sample identity (bottom left) and Louvain cluster distribution per sample, color-coded by cluster identity (bottom right). GBM molecular subtype analysis based on average gene expression of individual cells in each Louvain cluster for *PTEN*⁻/⁻;*NF1*⁻/⁻ (c) and *TP53*⁻/⁻;*PDGFRAΔB-9 (d) tumors. For *PTEN*⁻/⁻;*NF1*⁻/⁻ (e) and *TP53*⁻/⁻;*PDGFRAΔB-9 (f) tumors, stemness scores were calculated for each individual cell and results overlaid on a UMAP plot (top left) or summarized as violin plots for each cluster (top right). Cells were categorized based on their cell cycle status (G1, G2M or S), overlaid on a UMAP plot (bottom left) and their distribution was calculated for each Louvain cluster (bottom right).

Differentially expressed genes which, if stratified by sample, show no clear patterns, as would be expected when intra-tumor heterogeneity is present at the single-cell level (Supplementary Fig. 15). Furthermore, *PTEN*⁻/⁻;*NF1*⁻/⁻ iHGG secondary tumors appear more homogenous than their *TP53*⁻/⁻;*PDGFRAΔB-9 counterparts as each one of the triplicates is represented in all clusters. On the contrary, *TP53*⁻/⁻;*PDGFRAΔB-9 iHGG secondary tumors show clusters containing cells from all samples but also several clusters comprised of only two or almost exclusively one sample, indicating increased inter-tumor variability.
Significantly, when GBM molecular subtypes of each cluster are inspected, $PTEN^{−/−}$; $NF1^{−/−}$ samples show, for the most part, a homogenous mesenchymal signature, with the exception of two clusters (‘E’ and ‘G’) (Fig. 7c). In stark contrast, subtype signatures of each cluster derived from $TP53^{−/−}$; $PDGFRα^{Δ8−9}$ tumors show unique combinations, with three clusters accounting for the majority of the proneural signature and smaller contributions by the remaining clusters with the exception of cluster ‘O’ which has a clear mesenchymal signature (Fig. 7d). Notably, for both models, intra-tumor heterogeneity is also observed as each cluster presents signatures of different molecular subtypes, a hallmark of GBM patient samples. Finally, some $TP53^{−/−}$; $PDGFRα^{Δ8−9}$ tumor clusters also score higher for stemness in comparison to $PTEN^{−/−}$; $NF1^{−/−}$ clusters; while both models show a heterogenous composition of cycling and noncycling cells (Fig. 7e, f). In conclusion, our isogenic models faithfully recapitulate HGG pathobiology, including inter- and intra-tumor heterogeneity, differential drug sensitivity, ecDNA amplifications, and rapid clonal evolution. Variations of this tumor avatar platform can be applied to different types of cancers and will allow, amongst other things, a study of clonal evolution, longitudinal assessment in vitro and in vivo, and genotype-based therapeutic vulnerabilities deciphered in an isogenic background.

**Discussion**

We generated isogenic iHGG models from hiPSCs by introducing different combinations of genetic alterations characteristic of this disease. These models both recapitulated pathological features of HGG, but at the same time, displayed distinct mutation-dependent variation in histological morphology, gene expression, and ploidy. Some of our findings were consistent with previous mouse models. For example, $NF1$-deleted tumors and PDGF-driven tumors showed features of mesenchymal and proneural subtype, respectively in mouse models, where the former tumors were resistant to TMZ while the latter tumors were sensitive to the drug. Such consistency among our models, patient samples, and previous models suggests the key roles of small numbers of driver mutations in determining tumor phenotypes. The heterogeneity presented in our models is an essential feature for the proper modeling of GBM as this prominent characteristic is a confounding aspect making these tumors difficult to treat. Although there have been several models generated from pluripotent cells and genome editing technologies to date, our iHGG models are distinct in the way that they present exact pathognomonic features of GBM, and reproduce the heterogeneity of the disease from isogenic cells. These approaches have been applied for models of other brain tumors such as medulloblastomas. One limitation in our approach is that genome engineering was performed in hiPSCs, an irrelevant cell of origin for GBM. However, the fact that tumor models derived from appropriately differentiated NPCs from edited hiPSCs recapitulate GBM pathobiology suggests that our platform has potential for broader cancer modeling using various cell lineage differentiation protocols applied to hiPSCs. Furthermore, recent development of 3D in vitro brain tumor models using cerebral organoid suggests that several different combinations of oncogenic mutations give rise to expanding tumor-like components within organoids, which would approximate in vivo conditions better than conventional cell culture conditions. Such applications of advanced differentiation technology of pluripotent stem cells and genome engineering enabling introduction of genetic alterations actually observed in patient samples, could further develop the next generation of cancer models.

A combination of $EGFR$ activation and inactivation of $Ink4a/Arf$, which are common co-occurring genetic alterations observed in high-grade gliomas, has been shown to play a role in dedifferentiation of astrocytes through the process of gliomagenesis. How the genetic alterations we engineered in NPCs contribute to the formation of iHGGs is to be further studied, and presents an ideal platform to investigate mechanisms promoting transformation guided by cell lineage. Also, how the NPCs with GBM associated mutations, which prior to orthotopic engraftment do not show GBM subtype specific transcriptome signatures, present such signatures through the process of in vivo transformation, is to be further investigated.

Once xenograft tumors were obtained with our proneural and mesenchymal iHGG models, cells from these tumors maintained tumorigenicity and formed secondary tumors in vivo resembling the original tumors, as also seen in PDX models. Owing to this characteristic, the iHGG cells can be passaged and maintained as cell lines once tumors are obtained. Further, as shown here, it is quite feasible to introduce different combinations of genetic alterations in hiPSCs, and such different edits result in divergent phenotypes. Thus, by expanding our gene-editing spectrum, we expect that these models would enable us to evaluate the influence of selected driver genetic alterations found in different cancer types, which is less feasible with PDX models due to numerous acquired passenger mutations and genetic backgrounds that are highly variable sample to sample.

Another striking finding in our iHGG models was aneuploidy accompanied by ecDNA, observed in $TP53^{−/−}$; $PDGFRα^{Δ8−9}$ tumors. In our previous analyses, ecDNA was observed in more than 80% of PDX cell lines derived from GBM, suggesting that ecDNA formation is a fundamental feature in the pathogenesis of GBM. Interestingly, a mouse model with a combination of $CDKN2A^{−/−}$ and a $PDGFRα$ point mutation showed brain tumors with double minute chromosomes or ecDNA, while another $CDKN2A^{−/−}$ mouse model that generated brain tumors after irradiation similarly had ecDNA. Using human intestinal stem cells that were edited for the most commonly mutated colorectal cancer genes ($APC$, $TP53$, $KRAS$ and $SMAD4$), extensive aneuploidy occurred and these quadruple mutant cells grew as tumors in immunocompromised mice with features of invasive carcinoma. Together with our results, these previous models suggest alterations in $TP53$ or $CDKN2A$, which are commonly affected in GBM as well as other cancer types, play an essential role in the genesis of chromosomal instability that results in aneuploidy or ecDNA formation. In summary, we propose a modeling system for HGG by introducing different combinations of essential genetic alterations in hiPSCs, which result in tumor avatars faithfully recapitulating histology, gene expression signatures, and cytogenetic features of HGG. As these avatars are faithfully expressing gene expression signatures characteristic of GBMs, we expect that these models will be a useful platform to study cancer biology based on genetic drivers, cell of origin defined by the differentiation program of genome-edited iPSCs, and possible other parameters such as xenograft location and gender differences.

**Methods**

**Cell culture.** Experiments using human pluripotent stem cells were conducted under the regulations of the UCSD Human Research Protections Program, project number 151330XZ. Human iPS cells, CV-IPS-B cells were obtained from Dr. Lawrence S. B. Goldstein. CV-IPS-B cells were cultured on plates coated with Matrigel hESC-Qualified Matrix (Corning) in mTeSR1 media (Stemcell Technologies). NPCs were cultured on matrigel-coated plates in NPC maintenance media containing DMEM/F12 with GlutaMAX (Thermo Fisher Scientific), 1 x N-2 supplement (Thermo Fisher Scientific), 1 x B-27 supplement (Thermo Fisher Scientific), 50 mM ascorbic acid (Sigma), 3 μM CHIR99021 (Tocris) and 0.5 μM purmorphamine (Tocris). Spheres were cultured in suspension in DMEM/F12 with GlutaMAX (Thermo Fisher Scientific) with 1 x B-27 supplement, 20 ng/ml EGF (Stemcell Technologies) and 20 ng/ml bFGF (Stemcell Technologies).
Annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µl 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. Incubating 10 µM each of oligonucleotides, 1 × T4 DNA ligase buffer (New England Biolabs) containing cDNA obtained from 10 ng equivalent RNA were run on a CFX96 Real Time PCR instrument (Bio-Rad) to determine using 2-ΔΔCT formula. Primers used for the RT-qPCR are listed below.

| Primer Name | Sequence |
|-------------|----------|
| PDGFRA-intron 9-bottom | 5′-CCTGTTTTCCAGGGACTGAG-3′ |
| NF1-i31-f | 5′-TTGATGGTACATGACAAGGTGCGG-3′ |
| TP53-i1-f | 5′-CTCCACGGTACTCCTGTCTC-3′ |
| PDGFRAwt-RT-f | 5′-CACCGCTTTGGGGAGGTCTTTCGTC-3′ |
| PDGFRA-i9-r | 5′-GATACAAGGCTGTTAGAGAGATAATT-3′ |
| NF1-RT-f | 5′-CTCCACGGTACTCCTGTCTC-3′ |
| PTEN-RT-f | 5′-TTGATGGTACATGACAAGGTGCGG-3′ |
| NF1-RT-r | 5′-CTCCACGGTACTCCTGTCTC-3′ |
| PTEN-RT-r | 5′-TTGATGGTACATGACAAGGTGCGG-3′ |
| GAPDH-RT-f | 5′-CTCCACGGTACTCCTGTCTC-3′ |
| GAPDH-RT-r | 5′-TTGATGGTACATGACAAGGTGCGG-3′ |

Each of top and bottom oligonucleotides were phosphorylated and annealed by incubating 10 µM each of oligonucleotides, 1 x T4 DNA ligase buffer (New England Biolabs) at 37 °C for 30 min, 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. An annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µM annealed oligonucleotides, 1 x CutSmart buffer (New England Biolabs), 1 mM ATP (New England Biolabs) and 1 µl BBSI-HF (New England Biolabs) and 200U T4 ligase (New England Biolabs) at 37 °C for 5 minutes, 23 °C for 5 min for 30 cycles. Correct cloning of each sgRNA sequence was confirmed by Sanger sequencing using β6 sequencing primer: 5′-GATACAAGGCTGTTAGAGAGATAATT-3′.

Human iPSCs were cultured in 10 µL Y-27632 RH/ROCK pathway inhibitor for 2 hours before dissociation. The cells were dissociated to single cells using Accutase (Innovative Cell Technologies). The dissociated hiPSCs (1 × 10⁶ cells) were resuspended in 100 µl of supplemented solution of the Human Stem Cell Nucleofector 2b (Lonza). Electroporated hiPSCs were cultured on matrigel-coated Nucleofector Kit 1 (Lonza) containing 8 µg total of a combination of px458 plasmids targeting each gene and electroporated using B-016 program of Nucleofector 2, 1100 V, 900 ms. The PCR amplicons were visualized in agarose gels. Each of top and bottom oligonucleotides were phosphorylated and annealed by incubating 10 µM each of oligonucleotides, 1 x T4 DNA ligase buffer (New England Biolabs) at 37 °C for 30 min, 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. An annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µM annealed oligonucleotides, 1 x CutSmart buffer (New England Biolabs), 1 mM ATP (New England Biolabs) and 1 µl BBSI-HF (New England Biolabs) and 200U T4 ligase (New England Biolabs) at 37 °C for 5 minutes, 23 °C for 5 min for 30 cycles. Correct cloning of each sgRNA sequence was confirmed by Sanger sequencing using β6 sequencing primer: 5′-GATACAAGGCTGTTAGAGAGATAATT-3′.

Human iPSCs were cultured in 10 µL Y-27632 RH/ROCK pathway inhibitor for 2 hours before dissociation. The cells were dissociated to single cells using Accutase (Innovative Cell Technologies). The dissociated hiPSCs (1 × 10⁶ cells) were resuspended in 100 µl of supplemented solution of the Human Stem Cell Nucleofector 1 (Lonza) containing 8 µg total of a combination of px458 plasmids targeting each gene and electroporated using B-016 program of Nucleofector 2, 1100 V, 900 ms. The PCR amplicons were visualized in agarose gels. Each of top and bottom oligonucleotides were phosphorylated and annealed by incubating 10 µM each of oligonucleotides, 1 x T4 DNA ligase buffer (New England Biolabs) at 37 °C for 30 min, 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. An annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µM annealed oligonucleotides, 1 x CutSmart buffer (New England Biolabs), 1 mM ATP (New England Biolabs) and 1 µl BBSI-HF (New England Biolabs) and 200U T4 ligase (New England Biolabs) at 37 °C for 5 minutes, 23 °C for 5 min for 30 cycles. Correct cloning of each sgRNA sequence was confirmed by Sanger sequencing using β6 sequencing primer: 5′-GATACAAGGCTGTTAGAGAGATAATT-3′.

Human iPSCs were cultured in 10 µL Y-27632 RH/ROCK pathway inhibitor for 2 hours before dissociation. The cells were dissociated to single cells using Accutase (Innovative Cell Technologies). The dissociated hiPSCs (1 × 10⁶ cells) were resuspended in 100 µl of supplemented solution of the Human Stem Cell Nucleofector 1 (Lonza) containing 8 µg total of a combination of px458 plasmids targeting each gene and electroporated using B-016 program of Nucleofector 2, 1100 V, 900 ms. The PCR amplicons were visualized in agarose gels. Each of top and bottom oligonucleotides were phosphorylated and annealed by incubating 10 µM each of oligonucleotides, 1 x T4 DNA ligase buffer (New England Biolabs) at 37 °C for 30 min, 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. An annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µM annealed oligonucleotides, 1 x CutSmart buffer (New England Biolabs), 1 mM ATP (New England Biolabs) and 1 µl BBSI-HF (New England Biolabs) and 200U T4 ligase (New England Biolabs) at 37 °C for 5 minutes, 23 °C for 5 min for 30 cycles. Correct cloning of each sgRNA sequence was confirmed by Sanger sequencing using β6 sequencing primer: 5′-GATACAAGGCTGTTAGAGAGATAATT-3′.

Each of top and bottom oligonucleotides were phosphorylated and annealed by incubating 10 µM each of oligonucleotides, 1 x T4 DNA ligase buffer (New England Biolabs) at 37 °C for 30 min, 95 °C for 5 min and by cooling down to 25 °C at 0.1 °C/s using a thermocycler. An annealed oligonucleotides were cloned into px458 by incubating 25 ng px458, 1 µM annealed oligonucleotides, 1 x CutSmart buffer (New England Biolabs), 1 mM ATP (New England Biolabs) and 1 µl BBSI-HF (New England Biolabs) and 200U T4 ligase (New England Biolabs) at 37 °C for 5 minutes, 23 °C for 5 min for 30 cycles. Correct cloning of each sgRNA sequence was confirmed by Sanger sequencing using β6 sequencing primer: 5′-GATACAAGGCTGTTAGAGAGATAATT-3′.
Trypsin and D.Nase. Tumor tissue was resuspended by pipetting up and down several times through a glass Pasteur pipette. Dissociated tissue was filtered through a strainer and was spun down by centrifugation at 400 × g for 3 min. Cells were resuspended in 1 ml of PBS and 9 ml of ACK lysing buffer (Invitrogen) and were incubated at 37 °C for 10 min to remove red blood cells. Approximately 1 × 10^9 cells were resuspended in 100 ml of MACS/BSA buffer (Miltenyi Biotech) and were incubated with 2 µl of Fc blocking solution (BioLegend) for 5 min on ice. After blocking, 5 µl of PE-conjugated antihuman HLA-A,R,C antibody (BioLegend) was added and cells were incubated for 15 min on ice. Stained cells were washed twice with 500 µl of MACS/BSA buffer. PE-positive cells were then sorted using a flow cytometer (SH800, SONY). Sorted human iHGG cells were maintained in DMEM/F12 with GlutaMAX (Thermo Fisher Scientific) with 1 × B-27 supplement (Thermo Fisher Scientific), 20 ng/ml EGF (Stemcell Technologies) and 20 ng/ml bFGF (Stemcell Technologies).

**Extreme limiting dilution assay.** Extreme limiting dilution assay was performed based on a previous literature. In detail, NPCs and iHGG spheres were dissociated into single cells using accutase (Innovative Cell Technologies), washed with PBS, and resuspended at 2.5 × 10^5 cells in 2 µl PBS supplemented with 0.1% BSA per well. Cells were dissociated using accutase (Innovative Cell Technologies), PE-positive cells were then sorted on the BD FACSAria II (BD Bioscience) into a 96-well plate with five replicates for each experimental condition. The total number of spheres, per well and per treatment, were quantified after 14 days in culture. Data were analyzed by extreme limiting dilution analysis ([http://bioinf.wehi.edu.au/software/eldai/](http://bioinf.wehi.edu.au/software/eldai/)).

**Secondary tumor models and temozolomide treatment.** Primary iHGG spheres were dissociated using accutase (Innovative Cell Technologies), washed with PBS, and resuspended at 2.5 × 10^5 cells in 2 µl PBS supplemented with 0.1% BSA per well. Cells were dissociated using accutase (Innovative Cell Technologies), PE-positive cells were then sorted on the BD FACSAria II (BD Bioscience) into a 96-well plate with five replicates for each experimental condition. The total number of spheres, per well and per treatment, were quantified after 14 days in culture. Data were analyzed by extreme limiting dilution analysis ([http://bioinf.wehi.edu.au/software/eldai/](http://bioinf.wehi.edu.au/software/eldai/)).

**Cytogenetics.** Metaphase cells were obtained by treating cells with Karyomax (Gibco) at a final concentration of 0.1 µg/ml for 1–3 h. Cells were collected, washed in PBS, and resuspended in 0.075 M KCl for 15–30 min. Cells were kept on ice until hypotonicity was lost. After block, they were fixed with Carnoy’s mixture (3:1 methanol/glacial acetic acid) was added dropdown to stop the reaction. Cells were washed an additional three times with Carnoy’s fixative, before being dropped onto humidified glass slides for metaphase cell preparations. DAPI was added to the slides. Images were captured with an Olympus FX1000 confocal microscope.

**Spectral karyotyping analysis.** was performed at Applied Spectral Imaging. Genomic DNA extracted from NPCs and iHGG cells using DNeasy blood and tissue kit (Qiagen) was analyzed by digital karyotyping using Illumina HiScan system (Illumina).

**RNA sequencing.** Total RNA was assessed for quality using a Agilent Tapestation. For this article is available as a Supplementary Information. The RNA libraries were generated using Illumina’s TruSeq Stranded mRNA Sample Prep Kit (Illumina) following manufacturer’s instructions. RNA-seq reads were aligned to the human genome (hg19) with STAR 2.4.0 h (outFilterMultimapNmax 20, outFilterMismatchNmax 999, outFilterMismatchNoverLmax 0.04, outFilterMultimapNoverLmax 20, GlataMAX (Thermo Fisher Scientific) with 1 × B-27 supplement (Thermo Fisher Scientific), 20 ng/ml EGF (Stemcell Technologies) and 20 ng/ml bFGF (Stemcell Technologies). Z-scores were determined by subtracting the mean from each expression value and dividing by the standard deviation.

**scRNA-seq and analysis.** For the scRNA-seq of secondary tumor cells, the tumors were dissected from mouse brains, cut into small pieces, and then incubated in HBSS (Sigma) containing 1× trypsin (Sigma) at 37 °C for 20 min, followed by mechanical dissociation using glass pipettes to obtain single cells. Cultured sphere cells were dissociated using accutase (Innovative Cell Technologies). Single cells were processed through the Chromium Single-Cell Gene Expression Solution using the Chromium Single Cell 3‘ Gel Bead, Chip and Library Kits v2 (10× Genomics) as per the manufacturer’s protocol. In brief, single cells were resuspended in 0.04% BSA in PBS. Ten thousand total cells were added to each channel with an average recovery of 3040 cells. The cells were then partitioned into Gel Beads in Emulsion in an Illumina Chromium instrument, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification, shearing and 5’ adapter and sample index attachment. Agilent High Sensitivity D5000 Screen Tape Array (Agilent Technologies) was performed for QC of the libraries. Libraries were sequenced on an Illumina NovaSeq. De-multiplexing, alignment to the hg19 transcriptome and unique molecular identifier-collapsing were performed using the Cellranger toolkit (version 2.0.1) provided by 10× Genomics. A total of 42,558 cells with >53,000 mapped reads per cell were processed. Analysis of output digital gene expression matrices was performed using the Scapy v1.3.3 package. For single samples sequenced and analyzed, 10 samples were concatenated and all genes that were not detected in at least 20 single cells were discarded, leaving 20,521 genes for further analyses. Cells with fewer than 600 or more than 8000 expressed genes as well as cells with more than 80,000 transcripts or 0.1% mitochondrial expressed genes were removed from the analysis. For the different sample subset combination analysis filtering steps were the same with the exception of specific gene and transcript thresholds for which cells were removed: PTEN+/−;NF1+/−; tumors (fewer than 600 or more than 7000 expressed genes and more than 50,000 transcripts), PTEN+/−;NF1−/−; tumors (fewer than 600 or more than 8000 expressed genes and more than 80,000 transcripts), PT53−/−;PDGFRα−/−; tumors (fewer than 600 or more than 7000 expressed genes and more than 50,000 transcripts), PT53−/−;PDGFRα−/−; tumors (fewer than 600 or more than 8000 expressed genes and more than 80,000 transcripts), PDGFRα−/−; tumors, PDGFRα−/−; tumors, PDGFRα−/−; tumors, PDGFRα−/−; tumors, respectively). With these principal components, neighborhood graphs were computed with 20 neighbors and standard parameters with the pp.neighbors function. Louvain clusters were computed with the igLouvain function and standard parameters (and 0.4, 0.4, 0.7 and 0.4 resolution for PTEN+/−;NF1−/−; cells, PTEN−/−;NF1−/−; cells, PT53−/−;PDGFRα−/−; cells, respectively). Single-cell mean and expression mean per sample heatmaps were generated with the pl heatmap and pl matrixplot functions, respectively. Single-cell scores for TCGA molecular subtypes as well as stemness and cell cycle genes (see Supplementary Data 1) were computed with the t lucrasc_genes and t lucrasc_genes_cycle functions, respectively. Differentially expressed genes were determined for each set of Louvain clusters with the t lucrasc_group function (method = “wilcox”). For GO analysis of primary and secondary spheres, differentially expressed genes of each Louvain cluster with log2fold over 0.5 and p-adjusted values under 0.05 were used as inputs (see Supplementary Data 2) on Metascape (multigene list option and standard parameters), using all expressed genes for the 14 samples as background. Enrichment results are summarized in Supplementary Data 3.

**Statistical analyses.** All statistical analyses were performed using GraphPad Prism 6 software. Data are representative of results obtained in at least three independent experiments. Data sets were analyzed by unpaired t-test to determine significance (p < 0.05). Kaplan–Meier curves and comparison of survival were analyzed using Log-rank (Mantel–Cox) test.

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

**Data availability** Single-cell RNA sequencing and bulk RNA sequencing data have been deposited in the Gene Expression Omnibus database under GSE133479 and GSE133509, respectively. All the other data supporting the findings of this study are available within the article and its supplementary information files and from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file.

Received: 24 October 2019; Accepted: 20 December 2019; Published online: 28 January 2020
other authors, and supervised all aspects of the study. T.K. and I.A.C. contributed equally to this work. J.A.B. and Se.M. contributed equally to this work.

Competing interests
The authors declare the following competing interests: P.S.M. is a co-founder of Boundless Bio, Inc. (BB). He has equity interest in the company and serves as the chair of the Scientific Advisory Board. V.B. is a co-founder, serves on the scientific advisory board and has an equity interest in BB and Digital Proteomics, LLC (DP), and receives income from DP. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. BB and DP were not involved in the research presented here. K.M.T became an employer of Boundless Bio after submission of this manuscript. G.W.Y. is a co-founder, member of the Board of Directors, equity holder, and paid consultant for Eclipse BioInnovations. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. The remaining authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14312-1.

Correspondence and requests for materials should be addressed to G.W.Y. or F.B.F.