Comparative epigenetic analysis of tumour initiating cells and syngeneic EPSC-derived neural stem cells in glioblastoma

Claire Vinel1, Gabriel Rosser1,9, Loredana Guglielmi1,9, Myrianni Constantinou1,9, Nicola Pomella1, Xinyu Zhang1, James R. Boot1, Tania A. Jones1, Thomas O. Millner1, Anaelle A. Dumas1, Vardhman Rakyan1, Jeremy Rees2, Jamie L. Thompson3, Juho Vuononvirta4, Suchita Nadkarni4, Tedani El Assan2, Natasha Aley5, Yung-Yao Lin6,7, Pentao Liu7, Sven Nelander8, Denise Sheer1, Catherine L. R. Merry3, Federica Marelli-Berg4, Sebastian Brandner2,5 & Silvia Marino1

Epigenetic mechanisms which play an essential role in normal developmental processes, such as self-renewal and fate specification of neural stem cells (NSC) are also responsible for some of the changes in the glioblastoma (GBM) genome. Here we develop a strategy to compare the epigenetic and transcriptional make-up of primary GBM cells (GIC) with patient-matched expanded potential stem cell (EPSC)-derived NSC (iNSC). Using a comparative analysis of the transcriptome of syngeneic GIC/iNSC pairs, we identify a glycosaminoglycan (GAG)-mediated mechanism of recruitment of regulatory T cells (Tregs) in GBM. Integrated analysis of the transcriptome and DNA methylome of GBM cells identifies druggable target genes and patient-specific prediction of drug response in primary GIC cultures, which is validated in 3D and in vivo models. Taken together, we provide a proof of principle that this experimental pipeline has the potential to identify patient-specific disease mechanisms and druggable targets in GBM.

1Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, UK. 2Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, Queen Square, London, UK. 3Stem Cell Glycobiology Group, Biodiscovery Institute, University of Nottingham, Nottingham, UK. 4The William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, UK. 5Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, Queen Square, London, UK. 6Stem Cell Laboratory, National Bowel Research Centre, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, 2 Newark Street, London, UK. 7Faculty of Medicine, School of Biomedical Sciences, The University of Hong Kong, Hong Kong, Hong Kong. 8Department of Immunology Genetics and Pathology, Uppsala University, Uppsala, Sweden. 9These authors contributed equally: Gabriel Rosser, Loredana Guglielmi, Myrianni Constantinou. ✉Email: s.marino@qmul.ac.uk
glioblastoma (GBM) is the most common adult primary brain cancer; it cannot be completely resected by surgery and is resistant to conventional anticancer treatments, with a median survival of <15 months. Significant advances in understanding the genetics of these tumours have had, so far, only a modest impact on the therapeutic approach to the disease.

One of the main challenges in effectively tackling GBM is a remarkable heterogeneity at both intertumoural and intratumoural levels. Defined genetic alterations together with DNA methylation and transcriptional profiles have identified molecular subtypes which account for the intertumoural heterogeneity of GBM. The intratumoural heterogeneity though is determined not only by the presence of non-neoplastic cells, such as immune cells, brain cells, and vessels within the tumour ecosystem but also by substantial molecular and functional differences, as well as plasticity within the tumour cell population itself (recently reviewed in ref. 6). Cancer stem cells (CSC) have been shown to contribute significantly to the tumour component of the intratumoural heterogeneity and therefore to resistance to therapy and tumour recurrence in many cancers (recently reviewed in ref. 7). In GBM, CSC (called GBM stem cells (GSC) or GBM initiating cells (GIC)), exhibit properties of self-renewal and multilineage differentiation, similarly to neural stem cells (NSC) in brain development and homeostasis, and give rise to neoplasms recapitulating the tumour from which they were derived in orthotopic xenograft models. Single-cell transcriptional profiling of GBM has recently confirmed that GIC recapitulates a normal neurodevelopmental hierarchy with glial progenitor-like cells at its apex and that they contribute to drug resistance. The molecular and functional relationship between GIC and NSC is further corroborated by data in genetically engineered mouse models and functional validation of targets. These primary patient-derived cell lines retain the ability to generate tumours reflecting the heterogeneity of the parental tumour as shown in barcoding experiments followed by characterization of clonal expansion. They recapitulate to some degree the cellular states and plasticity found in GBM and core regulatory circuits identified in these cells are maintained in matched primary tumours. Epigenetic remodelling which locks GIC into an undifferentiated state upon prolonged culturing and the lack of interactions between these cells and various other cell types existing in vivo are known limitations of this experimental system.

Characterization of the molecular landscape of gliomas has highlighted a key role for epigenetic deregulation in these tumours. Methylation of the promoter of the DNA repair gene MGMT was the first example of an epigenetic regulation of the expression of a gene serving as a predictive biomarker of drug response in neuro-oncology. At a global DNA level, the methylome of brain tumours was recently utilized to develop a machine learning approach for a DNA methylation-based classification of brain tumours which enables standardization and improved precision of tumour diagnoses. Harnessing the expanding knowledge of epigenetic deregulation in brain tumours, particularly in those currently untreatable, to identify novel therapeutic approaches is an ambitious but now a realistic opportunity.

Characterization of epigenetic deregulation in GBM has so far relied on comparisons between tumour samples obtained from different patients or between GBM and progenitor cells derived from an unrelated donor or foetal brain. Here, we have harnessed state-of-the-art stem cell technologies and next-generation sequencing methods to compare the epigenetic and transcriptional profile of IDH-wildtype GBM with that of patient-matched normal expanded potential stem cell (EPSC)-derived NSC (iNSC). We show that syngeneic iNSC can be used as a proxy to identify disease-specific mechanisms and to predict drug response in vitro and in vivo systems.

Results

Establishment and characterization of patient-derived GIC and syngeneic iNSC lines. We established an experimental pipeline to derive GIC and patient-matched (syngeneic) iNSC from 10 patients with IDH-wildtype GBM (Fig. 1a and Table S1 for clinical and molecular data). hGIC lines were established according to standard protocols. Cases were included in the study where the molecular subgrouping, determined by DNA methylation arrays, was conserved between bulk tumour and hGIC (Fig. 1b) and where correlation analysis of gene expression revealed a good normalized correlation value between samples from the same patient (Fig. S1a). A match of copy number alterations between bulk tumour and corresponding GIC was confirmed by visual inspection of whole genome and chromosome-level plots (Fig. S1b). The tumour-initiating potential of all 10 GICs was confirmed in vivo by an intracerebral xenografting assay in NOD SCID mice (Fig. S1c).

Syngeneic fibroblast cultures (Fig. 1a and Fig. S2a) were derived from the dura mater of the same patient and reprogrammed to expanded potential stem cells (EPSC). Selected lines were characterized by immunostaining for OCT3/4, SOX2, NANOG, SSEA4 and TRA-1-60 (Fig. 1c, top). Hierarchical clustering of DNA methylation array data demonstrated that EPSC cluster with the reference “embryonic stem cells” (ESC) (Fig. S2c, d) from several different studies. To exclude significant epigenetic aberrations, which might have introduced bias into downstream analyses of the EPSC or cell lines derived therefrom, we studied the methylome and transcriptome of our EPSC relative to those of ESC lines in greater detail. We show that the number of differentially methylated regions (DMRs) detected between the EPSC lines and two reference ESC lines is similar to the number observed with published iPSC lines generated using different strategies (HipSci Initiative and) (Fig. S2e). This difference in the methylome is small in comparison to the 321 DMRs identified when the two reference ESC lines are compared with each other. Very limited overlap of DMRs is observed within our lines, with no hypermethylated and only 15 hypomethylated DMRs shared amongst all patients (Fig. S2f).

We then compared gene expression levels derived from the RNA sequencing data of the cell lines generated in this study, two reference ESC lines and one additional study of iPSC that also sequenced parental cell lines. Hierarchical clustering of these datasets again demonstrated that our EPSC lines cluster closely with ESC (Fig. S2g). Comparative expression analysis revealed 250 differentially expressed (DE) genes between the two reference ESC lines. The numbers of DE genes between EPSC with ESC are more variable than with the methylation profiles, with increased numbers of upregulated DE genes in two of our EPSC lines and one of the reference iPSC lines (Fig. S2h). Integration of the DNA methylation and transcriptomic datasets revealed a low level of association between DMR and DE genes: only 11 genes appear in both datasets and none of those are shared across all patient-derived cell lines (Table S2). This remarkable epigenetic similarity of our EPSC lines to ESC was corroborated by their trilineage differentiation capacity (ectoderm, endoderm, and mesoderm) in an embryoid body formation assay (Fig. 1c, bottom). EPSC were then induced to differentiate into iNSC using a commercially available protocol (denoted iNSC). Genome-wide expression...
and methylation analysis confirmed that iNSC\textsuperscript{Gibco} cluster together with various reference NSC lines and distinctly from iPSC and fibroblasts (Fig. 1d, e). These cells retained a normal karyotype (Fig. S2b), expressed NSC markers at protein level (Fig. S2i), formed neurospheres (Fig. S2j), downregulated the expression of pluripotency markers (Fig. S2k) and differentiated into all brain lineages in vitro (Fig. 1f). Importantly they homed to the subventricular zone (SVZ) upon intracerebral injection in adult mice (Fig. 1g).

In summary, we show that GIC and syngeneic iNSC pairs can be established from surplus surgical tumour material and a strip of dura mater, respectively.
iNSC are a suitable proxy for endogenous NSC to identify genes differentially expressed in a GBM mouse model. To assess the biological relevance of iNSC as a surrogate for endogenous NSC (eNSC) to study GBM biology, we derived eNSC from the SVZ of three adult mice, as well as fibroblasts from the dura mater of the same mice (Fig. 2a and Fig. S3a). Fibroblasts were reprogrammed into EPSC, as described above, which was confirmed by the expression of stemness markers at RNA and protein levels (Fig. S3b, c) as well as their ability to form embryoid bodies (EB) (Fig. S3d). These EPSC were then induced into NSC by the same protocol used in the human context above (denoted iNSC[ Gibco]) and also by a previously published customized protocol32 which had been validated in the mouse context (denoted iNSC[N2B27]). Comparative transcriptomic analysis clustered eNSC, iNSC[Gibco] and iNSC[N2B27] together close to reference NSC derived from mice of the same genetic background33,34 (C57Bl6, denoted rNSC[Syn]) but separately from a reference NSC derived from non-syngeneic mice35 (129sv/ev, etc.).
Fig. 2 EPSC-derived iNSC are a suitable proxy for eNSC in comparative studies with GIC in mice. a Schematic of experimental workflow. EPSC: expanded potential stem cells, iNSC: induced neural stem cells, eNSC: endogenous neural stem cells, mGIC: mouse glioblastoma initiating cells. b Two dimensional principal component analysis of cell types, representing normalized RNA sequencing data of eNSC, iNSC Gibco, iNSC N2B27, two rNSC syn, one rNSC non syn, two astrocytes, one neuron, and one GMP (granulocyte/monocyte progenitor). Astrocytes, neuron, and GMP are labelled as rOther syn. eNSC: endogenous/inferred/referenced neural stem cells, Gibco: EPSC neural induction with Gibco commercial protocol. N2B27: EPSC neural induction with a bespoke published protocol32, syn: syngeneic, non-syn: non-syngeneic. c Venn diagrams of overlapping genes (TPM > 1) in iNSC Gibco and eNSC, rNSC syn, rNSC non syn or neurons; results are expressed as percentages of genes shared or specific to each type of cells. d Three-dimensional principal component analysis of cell types, representing genome-wide methylation RRBS-Seq data (M-values) of eNSC, iNSC Gibco, Mueller cells, muscle satellite cells, hematopoietic stem cells (HSC), granulocyte-macrophage progenitor (GMP), and megakaryocyte/erythrocyte progenitors (MEP). e Two-dimensional density plots of DNA methylation profiles (M-values) between iNSC Gibco and eNSC (top) and between iNSC Gibco and muscle satellite cells (bottom). A median distribution of M-values has been obtained for each cell type across replicates before performing Spearman’s correlation. The R² value is shown for each plot. f Bar plots showing the relative abundance of genomic regions (distal, exon, intron, promoter, 3' UTR, 5' UTR) in methylated (top) and unmethylated (bottom) CpG sites across all cell types. g Left: combined analysis of expression profile and methylation levels of eNSC and iNSC (1107 genes are plotted). Genes are considered expressed if log (TPM) > 0 and methylated if M > 0. Concordant genes are depicted in orange and blue, according to the orientation. Right: Venn diagrams showing the overlap of concordant genes between eNSC and iNSC Gibco for both orientations. h Coexpression of SOX2 (red) and NESTIN (green) in eNSC, iNSC Gibco, and iNSC N2B27 with DAPI nuclear counterstaining (representative images, n = 3 for each cell type). The scale bar is 20 μm. i Percentage of genes classified as Differentially Expressed (DE) and not DE in comparisons between GIC vs NSC39 or mGIC vs eNSC, iNSC Gibco, rNSC syn, and rNSC non syn respectively.

denoted rNSC non syn) as well as other brain (astrocytes and neurons) and non-brain (monocytes) cells derived from syngeneic mice36–38 (C57Bl6, denoted rOther syn) (Fig. 2b and Fig. S4). In particular, the driver of clusters, the principal component PC1, clearly separates eNSC, iNSC, and rNSC non syn from the other cell types and is not able to discriminate eNSC from iNSC and the syngeneic reference (Fig. 2b). In agreement with these results, the highest overlap of expressed genes was noted between eNSC and rNSC syn (88.8%), followed by iNSC Gibco and iNSC N2B27 (85.4%) and eNSC and iNSC Gibco (80.5%) (Fig. 2c and Fig. S3e). The degree of overlap reduced to 69.4% when iNSC Gibco were compared with non-syngeneic reference NSC (Fig. 2c), and to 64.7 and 61.5% when eNSC and iNSC Gibco were compared to syngeneic neurons, respectively (Fig. S3e).

Comparative analysis of 16S RNA methylation profiles between eNSC, iNSC Gibco and syngeneic non-brain cells shows clustering of eNSC and iNSC Gibco (Fig. 2d) and higher correlation between iNSC Gibco and eNSC than the other cell types, including stem cells of mesenchymal origin (Fig. 2e and Fig. S3f). We compared the proportions of the annotated regions between cell types and showed that the probability of those proportions being different is lowest between eNSC and iNSC Gibco, highlighting a higher degree of similarity compared to the other cell types (Fig. 2f and Fig. S3g). Additionally, comparative analysis of combined methylene and transcriptome of eNSC and iNSC Gibco confirmed a high degree of overlap (82%) of concordant genes (methylated and low expression and vice-versa) (Fig. 2g).

Phenotypic and functional characterization of eNSC, iNSC Gibco, and iNSC N2B27 confirmed coexpression of the NSC markers Nestin and Sox2 (Fig. 2h) and their ability to form neurospheres (Fig. S3h).

Finally, we queried the value of these iNSC in identifying DE genes in a glioblastoma mouse model39. We show that syngeneic iNSC Gibco identified more DE genes as compared to non-syngeneic NSC (rNSC non syn), 70% as compared 50% (Fig. 2i).

Our results in mouse models show that iNSC are reasonably identified in the genes DE in hGIC relative to iNSC (Fig. S6a, c). A significant correlation was noted between enrichment for pathways linked to the biosynthesis of GAGs, linear polysaccharides including HS and DS, in the hGIC/iNSC comparison, and higher proportion of regulatory T-cells (Tregs) as well as a lower proportion of macrophages in the corresponding bulk tumour (Fig. S6c and Fig. 3b). This correlation was not identified when a non-syngeneic iNSC comparator was used for the identification of hGIC DE genes (Fig. S6d). Immunostaining for the Tregs marker FoxP3 and CD4 in the corresponding bulk tissue samples confirmed a higher number of Tregs among the CD4+ tumour infiltrating lymphocytes (TIL) (Fig. 3c), which was most prominent in those patients predicted to have a higher proportion of Tregs as assessed by CD4/DS biosynthetic pathway deregulation (Fig. S7a). Increased levels of CS in GIC as compared to the matched iNSC were confirmed using immunocytochemistry and were neutralized by targeted digestion using chondroitinase ABC enzyme (chABC) (Fig. S7b). We
hypothesized that deregulated GAG biosynthesis may influence Tregs infiltration in a proportion of GBM patients, which contributes to the generation of an immunosuppressed microenvironment in GBM [41]. Biosynthesis of CS/DS involves the activity of multiple sulfotransferases and other synthetic enzymes. To read out this activity, we analyzed the composition of GAGs synthesized by iNSC and GIC cultures. GAGs were isolated from GIC and iNSC cell extracts and conditioned media, digested to disaccharides, and analyzed by RP-HPLC (Fig. 3d and Fig. S7c). We observed an increase in 6-O-sulphation (D0a6) in iNSC compared to GIC, with a correspondingly lower proportion of HexA(2S)-GalNAc (D2a0) and HexA-GalNAc(4S) (D0a4) disaccharides (Fig. 3d). A similar pattern was observed in both cell/matrix derived CS/DS and CS/DS isolated from the conditioned
Fig. 3 Comparative transcriptome analysis of GIC/iNSC identifies a GAG-mediated mechanism of Tregs migration in a proportion of GBM. a Network representation of the enriched pathways identified in the 10 syngeneic GIC/iNSC comparisons. Nodes represent pathways; pie charts are shaded according to the patients in which the pathway is identified (see legend); size is proportional to the mean value of $-\log_{10}(FDR)$ across all patients. Edges represent deregulated genes shared between pathways (pooled over relevant patients); thickness is proportional to the number of genes shared. b Heatmap showing the Spearman rank correlation across patients between pathway enrichment in the GIC culture relative to matched iNSC, expressed as $-\log_{10}(p)$, and estimated cell type composition in the corresponding GBM bulk tissue sample. Cells with statistically significant correlation coefficient are highlighted ($p < 0.05$). c Representative histology of glioblastoma bulk tissue (Hematoxylin & Eosin (H&E) staining, top), immunostaining for CD4 (bottom left) and FoxP3 (bottom right). FoxP3/CD4 ratio in all GBM samples (right). Each case has been stained once. Scale bar is 250 µm. d Disaccharide composition of CS isolated from GIC (black bars) and iNSC (grey bars) cells (plain bars) and from conditioned media (hatched bars) analyzed by RP-HPLC. D0a6: HexA-GalNAc; D0a6: HexA-GalNAc(6S); D2a0 HexA(2S)-GalNAc; D0a4: HexA-GalNAc(4S); D2a6 HexA(2S)-GalNAc(6S); D0a10: HexA-GalNAc(4S, 6S). Results are an average of two GIC/iNSC pairs, 19 and 31, (n = 2, two way ANOVA). (D0a6: p value (GIC vs iNSC cells/matrix) < 0.0001, p value (GIC vs iNSC CM) = 0.0009; D2a0/D0a4: p value (GIC vs iNSC cells/matrix) < 0.0001, p value (GIC vs iNSC CM) = 0.0003). Percentages of migrated immune cells in a transwell assay in the presence of GIC (black bars), their 24 h conditioned-media (CM, hatched black bars), iNSC (grey bars) and media only (light grey bars) after 4 h incubation. Percentage of Tregs (CD4$^{+}$CD8$^{−}$FoxP3$^{+}$CD25$^{+}$) (Fig. S7c). HS from GIC conditioned media had increased levels of 2-O-sulfation and 6-O-sulfation (D250 and D2S6) compared to iNSC conditioned media.

Next, we set out to assess the functional impact of the observed shift in GAG sulfation on Tregs by means of in vitro chemotaxis experiments. We showed increased migration of Tregs in the presence of GIC (black bars) compared to their matched-iNSC (dark grey bars) (Fig. 3e and Fig. S7d), an effect which was predominantly mediated by the cellular fraction as it was reduced when PBMC were incubated with GIC-conditioned media alone (hatched bars). This effect of GIC on Tregs migration was dependent on CS/DS as pretreatment with chABC, catalyzing digestion of CS and DS chains (Fig. 3e), blocked Tregs migration in the transwell assay (black bar with orange border). The observed GAG dependent Tregs migration is specific to this cell population and does not impact CD4$^{+}$FoxP3$^{+}$CD25$^{−}$ T cells (Fig. 3f and Fig. S7e), CD8$^{+}$FoxP3$^{−}$ (Fig. 3f and Fig. S7f) and monocytes (Fig. 3f and Fig. S7g) or NK cells (Fig. 3f and Fig. S7h).

In summary, differential transcriptomic analysis of GIC/iNSC syngeneic pairs identified a shift in GAG sulfation in a proportion of patients, which controls Tregs migration in GBM.

Integrated transcriptome and methylation analysis of syngeneic hGIC/iNSC pairs identifies patient-specific druggable targets. To explore the potential suitability of the platform to identify genes for which compounds exist, and which could be tested for patient selective efficacy, we set out to comparatively interrogate transcriptome and methylation profiles (Fig. 4a). The process yielded around 17,000 unique (gene, methylation cluster) pairs corresponding to around 12,000 unique genes (the background). We identified 2,222 unique differentially expressed pairs with a corresponding differentially methylated region in at least one patient, of which 1,565 (70%) were concordant, corresponding to 1,274 unique genes (Supplementary Data 1). Of the concordant pairs, 733 (47%) were specific to one patient corresponding to 639 distinct genes, ranging from 20 pairs in patient 50–177 in patient 18 (Fig. S8a). In contrast, a very small proportion (17 pairs, 1%) was shared across all patients. Furthermore, no single combination of patients shared a substantial number of DE/DMR pairs: all subsets had fewer than 25 (Fig. S8a). This emphasizes the high inter-patient heterogeneity of glioblastoma and the importance of a patient-specific approach.

We queried the drug-gene interaction database (DGIdb) against the 1,274 gene list and identified 1,920 distinct drug compounds known to interact with 237 of them (19%) (Supplementary Data 1). This is a highly significant enrichment relative to the <3% of background genes with known interactions (Fisher’s test $p < 10^{-6}$), confirming that our analytical pipeline is an effective approach to enrich for druggable targets. 645 (34%) of the compounds are predicted in a single patient, ranging from 15 in patient 17 to 205 in patient 18, further highlighting the importance of a patient-specific approach (Fig. S8b). This corresponds to 639 unique genes, of which 118 (18%) have matching drug compounds and are therefore candidates for potential follow-up (Supplementary Data 1).

In order to further prioritize targets, we shortlisted a subset with the even more stringent criterion that genes must be patient-specific at the level of DE and DMR independently (Fig. 4a). While genes in the longest may be DE or DMR in multiple patients providing that only one has the combination of both DE and DMR, genes in the shortlist are patient-specific independently at the level of DE and DMR. This approach yields 53 unique genes, of which 15 (28%) have matching drug compounds. Taking into account the mechanism of action (an agonist for a gene hypermethylated/downregulated and an antagonist for a gene hypomethylated/overexpressed), commercial availability of the compound, and their pharmacological properties (specificity, potential for blood barrier crossing), PGE1-OH targeting PTGER4 (patient 18) and Disulfiram (DSF) targeting ALDH3B1 (patient 19) were selected for experimental validation. NTRK2, identified in patient 30, was also taken forward to validation as it was predicted to be effectively targetable with Cyclostratin-B (CTX-B). Interrogation of a publicly available collection of 71 GICs with DNA methylome
Inhibitor uniquely in GIC that maintains a pool of slow-cycling (Fig. 4c). It has been suggested that it serves as a proliferation mediated regulation of PTGER4 expression.

Prostaglandin E1 alcohol (PGE1-OH) is a specific EP4 agonist that significantly reduces viability in GIC18 compared to the syngeneic iNSC (Fig. 5a, top), while it had a similar effect on GIC30 and iNSC30 (Fig. 5a, bottom). For patient 26, both GIC and iNSC did not respond to treatment with PGE1-OH but patient 30 (low expression of PTGER4 in both GIC and iNSC and no concordant differential methylation, both predicted to respond to PGE1-OH) had a similar effect on GIC30 and iNSC30 (Fig. 5a, bottom). For patient 26, both GIC and iNSC did not respond to treatment with PGE1-OH (Fig. 5g), in keeping with the in silico prediction.

Next, we used a catalytically inactive Cas9 (dCas9) fused with Tet1 (ten-eleven translocation Methylosine Dioxygenase 1) to assess the functional relevance of PTGER4 hypermethylation by site-specific DNA methylation editing. We show successful demethylation of an hypermethylated CpG dinucleotide within a CpG island in the PTGER4 promoter (cg04727116, Illumina Infinium Methylation EPIC) in GIC18 (GIC18 TET1/sgRNA4) (Fig. 5c and Fig. S10a) which resulted in increased expression of PTGER4 (Fig. 5b). GIC18TET1/sgRNA4 showed reduced proliferation compared to non-edited cells (Fig. 5d, black bars vs hatched black bars) and mice bearing a GIC18TET1/pgRNA (Fig. 5e). Importantly, no sensitivity to PGE1-OH treatment was seen in GIC18TET1/sgRNA4 derived xenograft surviving longer as compared to mice xenografted with GIC18TET1/pgRNA (Fig. 5e).

Patient-specific efficacy of Prostaglandin E1 alcohol in 2D and 3D syngeneic GLICO models is predicted by methylation-mediated regulation of PTGER4 expression. PTGER4 is the gene encoding the receptor for prostaglandin E2 (PGE2) and it belongs to the G-protein coupled receptor family. It is hypermethylated with a concordant, strongly significant down-regulation in GIC18 as compared to the syngeneic iNSC (Fig. 4b), and reduced expression of the gene was confirmed at protein level (Fig. 4c). It has been suggested that it serves as a proliferation inhibitor uniquely in GIC that maintains a pool of slow-cycling cells refractory to conventional therapy. In agreement with this hypothesis, higher expression of PTGER4 is observed in the tumour bulk of GBM18 and in the TCGA GBM cohort (Fig. S9a) as compared to GIC18, and a negative correlation was observed between the expression of PTGER4 and PROM1 (CD133), a GIC marker, in the TCGA datasets (Fig. S9b). Moreover, genetic overexpression of PTGER4 in GIC18 (Fig. S9c) negatively impacted the proliferation of the cells (Fig. S9d), in keeping with the interpretation that it has a tumour suppressive role in GIC. Prostaglandin E1 alcohol (PGE1-OH) is a specific EP4 agonist used clinically as a non-irritant bronchodilator. To assess whether the in silico prediction of its efficacy in patient 18 could be validated in vitro, we used patient 18 (low expression and hypermethylation in GIC as compared to iNSC, where only the GIC are predicted to respond to PGE1-OH), patient 30 (low expression of PTGER4 in both GIC and iNSC and no concordant differential methylation, both predicted to respond to PGE1-OH) and patient 26 (high expression of PTGER4 in both GIC and iNSC and no concordant differential methylation, both predicted not to respond to PGE1-OH) (Fig. S9e and Table S3). We show that treatment with PGE1-OH significantly reduces viability in GIC18 without affecting the matched iNSC (Fig. 5a, top), while it had a similar effect on GIC30 and iNSC30 (Fig. 5a, bottom). For patient 26, both GIC and iNSC did not respond to treatment with PGE1-OH (Fig. 5g), in keeping with the in silico prediction.

PGE1-OH treatment decreases proliferation specifically in GIC18 (Fig. 5b), without triggering toxicity (Fig. S9g) or apoptosis (Fig. S9h) and without affecting tumour sphere formation (Fig. S9i).
PGE1-OH is administered to patients via inhalation\textsuperscript{15}, but we are not aware of any other use of either PGE1-OH or other PTGER4 agonists, hence making in vivo validation not viable. Therefore, to provide further support to the selective efficacy of the drug in patient 18, we used the 3D cerebral organoid glioma (GLICO) model whereby GIC are grown within a cerebral organoid (CO), which can be used as an effective predictor of drug response in vivo\textsuperscript{47,48}. Taking advantage of the availability of patient-matched EPSC we set up a modified GLICO model where GFP-labelled GIC are co-cultured with CO derived from the matched EPSC (SYNGLICO, Fig. 5f). CO developed successfully, as demonstrated by multiple neural rosettes observed throughout the CO on morphology, with neuroectodermal cells, expressing SOX2 and NESTIN, arranged in radial fashion within the rosettes and NeuN-positive cells populating the neuropil between the neural rosettes (Fig. S10c). To assess the response to PGE1-OH, 100,000 GIC\textsuperscript{18GFP} and GIC\textsuperscript{30GFP} were incubated with 30 days old EPSC\textsuperscript{18-derived} and EPSC\textsuperscript{30-derived} CO respectively, for 5–7 days and treated with 50 and 100 µM PGE1-OH for 4 days. We show no significant difference in the overall viability of the co-cultures (Fig. S10d, e), while significantly decreased mean fluorescent intensity (MFI) of GIC\textsuperscript{18GFP} but not GIC\textsuperscript{30GFP} was
observed (Fig. 5g and Fig. S10e). Assessment of proliferation (Ki67 immunostaining) and apoptosis (cleaved Caspase 3 immunostaining) in CO + GIC18<sup>GFP</sup> treated with 50 µM PGE1-OH confirmed reduced proliferation and increased apoptosis in GIC18<sup>GFP</sup> (Fig. 5h and S11a).

Finally, we confirmed in our SYNLICO model that PGE1-OH effect was dependent on PTGER4 methylation as PGE1-OH treatment had no effect on CO + GIC18<sup>TET1/sgRNA4</sup> (Fig. 5i and S11b, c).

Our results show that PTGER4 functions as a tumour suppressor in selected GBM patients and regulation of its expression by DNA methylation predict response to PTGER4 activation therapy in 2D and 3D in vitro models.

**Patient-specific efficacy of disulfiram and cyclotraxin-B in vitro and in vivo.** Aldehyde dehydrogenase 3 Family Member B1 (ALDH3B1), a member of the aldehyde dehydrogenases family protecting against oxidative stress<sup>49</sup>, is hypomethylated specifically in GIC19 as compared to the syngeneic iNSC19 within our cohort (Fig. 4b). A concomitant upregulation of expression (Fig. 4b) was observed, which was confirmed at protein level (Fig. 4c). DSF, a specific inhibitor of the ALDH superfamily<sup>50</sup>, is known to reduce cell viability, proliferation, and self-renewal capacity, as well as to induce cell cycle arrest and apoptosis in paediatric brain tumours<sup>51</sup>.

We show in 2D monolayer cultures that DSF reduces cell viability (Fig. 6a, top) and induces cytoxicity (Fig. S12a, top) in GIC19 without eliciting an effect on the syngeneic iNSC. Importantly, GIC31 and iNSC31 which did not show differential expression or methylation for ALDH3B1, showed no sensitivity to the drug (Fig. 6a, bottom, and Fig. S12a, bottom). To further dissect the impact of DSF treatment on cellular functions, we assessed proliferation, tumour sphere formation as well as apoptosis and showed decreased proliferation (Fig. 6b) and tumour spheres formation (Fig. 6c and Fig. S12c) whilst apoptosis was increased (Fig. S12b) in GIC19 but not in GIC31. Concomitantly, total ALDH activity was significantly more impacted by DSF treatment in GIC19 as compared to GIC31 (Fig. S12d), suggesting the effect of DSF was specifically on ALDH in our experimental setting. Silencing of ALDH3B1 (Fig. S12e), which did not affect the expression of other ALDH isoforms (Fig. S12f), inhibited GIC19 proliferation (Fig. S12g) and confirmed that the impaired viability observed after DSF treatment is ALDH3B1-dependent as it is lost in GIC upon ALDH3B1 silencing (Fig. S12h).

These results were confirmed in the 3D SYNLICO model where DSF treatment impacted the MFI of GIC19<sup>GFP</sup>+ cells within CO + GIC19<sup>GFP</sup> without affecting viability (Fig. 6d, e and Fig. S13, left) indicating loss of GFP<sup>+</sup> tumour cells whereas no such effect was observed in GIC18<sup>GFP</sup>+ CO (Fig. 6d, e and Fig. S13, right). Reduced proliferation and increased apoptosis were confirmed in GIC19<sup>GFP</sup> cultured within CO19 treated with 10 and 50 µM DSF by immunostaining for Ki67 and cleaved-caspase-3 respectively (Fig. 6f, g).

In vivo experiments whereby xenografts were induced by intracerebral injection of GIC confirmed extended survival (Fig. 6h, left) of mice bearing a GIC19-derived xenograft upon treatment with DSF, with no effect in mice xenografted with GIC31 (Fig. 6h, right).

NTRK2 is a member of the neurotrophic tyrosine receptor kinase (NTRK) family regulating neuronal differentiation, as well as growth and survival (reviewed in ref. 52). Somatic alterations of this gene have been found in brain tumours including gliomas<sup>53</sup> and the expression of this gene is prognostically relevant in these tumours<sup>54</sup>. Recently, it has been shown to play a key role in the reciprocal signalling between GICs and their differentiated glioblastoma cell progeny<sup>55</sup>. TRK inhibition has emerged as an important therapeutic target in various cancers<sup>56</sup>. We show that NTRK2 is hypomethylated with a concordant upregulation in patient 30 (Fig. 4b), the latter validated at protein level (Fig. 4c).

Treatment with CTX-B, a selective, non-competitive antagonist of NTRK2, known to cross the BBB impacts on the viability of GIC with no effect on the syngeneic iNSC in patient 30 (Fig. 7a, top), while it has no effect on both GIC and iNSC from patient 18, where the molecular signature is not present (Fig. 6a, bottom). CTX-B treatment decreased proliferation of GIC 30, without affecting GIC18 (Fig. 7b) tumour sphere formation (Fig. S4a), cytoxicity (Fig. S4b), and apoptosis (Fig. S4c) were not affected. Silencing of NTRK2 (Fig. S14d) inhibits GIC30 proliferation (Fig. S14e) and confirms that the impaired viability observed after CTX-B treatment is NTRK2-dependent as it is lost in GIC upon NTRK2 silencing (Fig. S14f).
These results were confirmed in the 3D SYNLICO model where CTX-B treatment impacted the MFI of GIC30\(^{GFP}\) cells within EPSC30-derived CO without affecting viability, indicating loss of GFP\(^{+}\) GIC cells, with no significant effect on the MFI of GFP in GIC18\(^{GFP}\) grown within an EPSC18-derived CO (Fig. 7c, d and Fig. S15a).

In vivo experiments showed extended survival (Fig. 7e, left) in mice bearing GIC30-derived xenograft upon treatment with tat-Cyclotraxin-B\(^{57}\), while it had no effect in mice bearing GIC18-derived xenograft (Fig. 7d, right).

Our results show that deregulation of ALDH3B1 and NTRK2 specifically in GIC19 and GIC30 respectively, played a key pathogenetic role specifically in these patients and can be effectively targeted therapeutically in pre-clinical models.

Absence of overt gastrointestinal toxicity of the compounds at the GIC effective dose as assessed on intestinal organoids. Drug toxicity on the gastrointestinal tract is a known limiting factor of adjuvant chemotherapy in cancer\(^{58}\). Here, we have taken
advantage of the availability of patient-matched EPSC to generate intestinal organoids, which have been previously used to assess drug response\textsuperscript{59}. Expression of \textbeta-CATENIN, E-CADHERIN, VILLIN, and CDX2 confirmed successful differentiation (Fig. S16a). Treatment with PGE1-OH (50 and 100 \textmu M), DSF (10 and 50 \textmu M) or CTX-B (50 and 100 \textmu M) of the intestinal organoids did not show a significant reduction in proliferation or increase in apoptosis, in keeping with lack of measurable toxicity at the doses effective in GIC on a patient-specific basis (Fig. S16b–d).

**Discussion**

Here we describe an approach to generate syngeneic neural stem cells and use them as patient-specific comparator to glioma initiating cells for the discovery of pathogenetically relevant mechanisms in glioblastoma and to identify druggable targets in patients.

There is robust published evidence, both in mouse models\textsuperscript{60} and in human patients\textsuperscript{12}, that NSC are the cell of origin of at least a proportion of GBM. Deregulated transcriptional programmes and epigenetic mechanisms have been convincingly shown to play a role, together with the neoplastic transformation of these cells into GIC and in conferring essential tumour maintenance properties to the latter\textsuperscript{61}. INSC derived from iPSC have been extensively characterized and shown to share marker expression and functional properties with endogenous NSC\textsuperscript{61,62}. Here, we demonstrate that iNSC are remarkably similar to eNSC derived from the brain of the same animal and that they are superior to non-syngeneic eNSC in identifying GBM-defining expression in a mouse model. The strength of our study is the application of this concept to human GBM in a cohort of 10 patients to compare for the first time the epigenetic and transcriptional make-up of GIC with that of patient-matched normal iNSC. This comparison was not feasible to date, as patient-matched endogenous NSC are not surgically accessible and all epigenetic studies in GBM had so far compared epigenetic changes of GBM cells with each other or to comparators obtained from fetal brains\textsuperscript{44} or an unrelated donor\textsuperscript{35}.

Using this syngeneic comparative analysis, we reveal a mechanism of regulatory T cell migration in GBM, mediated via deregulation of GAGs in GIC, in a proportion of GBM patients. A decreased cellular immunity is characteristic of GBM, to which an increased proportion of Tregs among CD4+ TIL (tumour-infiltrating lymphocytes) has been shown to contribute both in patients\textsuperscript{63,64} and in pre-clinical models\textsuperscript{65,66}. Microenvironment-derived cytokines, such as CCL-2, IDO, and TIM4 (reviewed in\textsuperscript{67}) are responsible for the accumulation of Tregs in GBM; however, a role for tumour cells or GIC in modulating Tregs migration is less clear and a link between GAG-GIC and Tregs migration has not been previously described. Sulfation of GAGs plays a regulatory function in modulating extracellular signals in cell–cell and cell–matrix interactions during development and in pathological processes (reviewed in\textsuperscript{68}), such as cancer. Chondroitin sulfate proteoglycans (CSPGs) and their GAG side chains are key constituents of the brain extracellular matrix (ECM) implicated in promoting GBM invasion\textsuperscript{69}. Changes in GAG sulfation have been proposed as cancer biomarkers, however, person-to-person variation in GAG composition has hampered their applicability\textsuperscript{70}. Notably, our results show that GIC regulate Tregs migration specifically through a CS/DS-dependent mechanism both in vitro and in vivo and suggest that this is mediated by their altered sulfation both in secreted and membrane GAG from GIC. Therapeutic approaches aiming at modulating GAG sulfation, for example, to deplete infiltrating Tregs, may therefore require this type of patient-specific approach and the value of a FoxP3/CD4 index as a predictive biomarker in surgical pathology should be further assessed.

Importantly, we provide proof of principle that combined comparative analysis of the methylene and transcriptome of GIC and matched INSC identifies druggable targets in a patient-specific fashion in a proportion of cases. The ALDH inhibitor disulfiram, which has been trialled clinically in GBM patients with unmethylated MGMT because of its MGMT inhibitory role\textsuperscript{71}, may be more effective or have additional effects on patients with ALDH3B1-specific deregulation. Hypomethylation/high expression of NTRK2, a druggable tyrosine receptor kinase\textsuperscript{56} playing a role in the reciprocal signalling between GIC and their differentiated glioblastoma cell progeny\textsuperscript{55}, predicts response to the specific inhibitor cyclotraxin-B. Finally, we provide compelling evidence that PTGER4 agonists are effective in a patient-specific fashion predicted by hypermethylation/low expression of the gene in the syngeneic GIC/INSC comparison in 2D cultures and in a 3D syngeneic GLICO model. Development of compounds that can be administered in vivo will be required for further pre-clinical assessment of its applicability and assessment.

---

**Fig. 6 Comparative analysis of transcriptome and methylene of GIC/INSC identifies ALDH3B1/DSF as patient-specific drug sensitivity.** a Drug treatment of GIC (black curves) and INSC (grey curves) of patient 19 (top) and 31 (bottom) with doses ranging from 1 nM to 10 \mu M of DSF (orange dots) or vehicle (black dots). Results are expressed in percentage of viability on the vehicle and were measured at end point, area under the curve (AUC) was calculated from percentages of viability (n = 6 experiment repetitions, two tailed t-test, p value (patient 19) = 0.0155, p value (patient 31) = 0.3459). b Proliferation assay of GIC19 (top) and 31 (bottom) treated with vehicle (black bars) or DSF (orange hatched bars) as assessed by cell number after 4 days of culture. Results expressed as fold change of cell number at day 4 on day 1 of treatment and standardized on the vehicle treatment conditions (n = 3 experiment repetitions, two-tailed t-test, p value (patient 19) = 0.0012, p value (patient 31) = 0.2743). c Tumour spheres (TS) were counted after 4 days of treatment of GIC19 (top) and 31 (bottom) with vehicle (black bars) or DSF (orange hatched pattern bars) in non-adherent culture conditions. Results are an average of 10 \times 10 pictures (n = 3 experiment repetitions, two-tailed t-test, p value (patient 19) = 0.0273, p value (patient 31) = 0.7780). d, e Viability of total cells in SYNGLICO (Cerebral Organoids (CO)-iGIC\textsuperscript{SP}) as assessed as percentages of negative cells for the Zombie NIRTM dye after 4 days of treatment with vehicle or DSF (10 \mu M, 50 \mu M) (d) and GFP Mean Fluorescence Intensity (MFI) of GIC\textsuperscript{GPSP}+ cells (e). COGIC18: n = 12, GOIC19: n = 8, DSF 10 and 50 \mu M: n = 9 5 SYNGLICO per group, one way ANOVA; d) Patient 19: p value (Vehicle vs DSF 10 \mu M) = 0.0166, p value (Vehicle vs DSF 50 \mu M) = 0.0029; e) Patient 19: p value (Vehicle vs DSF 50 \mu M) = 0.4178. Representative GFP and Ki67 (f) or cleaved-caspase3 (g) immunofluorescence of SYNGLICO after 4 days of treatment with vehicle or DSF (10 \mu M or 50 \mu M). The scale bar is 10 \mu m. Quantification of double positive cells GFP + Ki67 (f) and GFP + cCASP3 (g) is shown on the right (n = 2 SYNGLICOs, GFP + Ki67: Vehicle: n = 13 fields, DSF 10 \mu M: n = 9 fields, DSF 50 \mu M: n = 11 fields, GFP + cCasp3: Vehicle: n = 14 fields, DSF 10 \mu M: n = 12 fields, DSF 50 \mu M: n = 10 fields, one-way ANOVA; f, p value (Vehicle vs DSF 10 \mu M) = 0.0033, p value (Vehicle vs DSF 50 \mu M) = 0.0004; g, p value (Vehicle vs DSF 10 \mu M) = 0.0006, p value (Vehicle vs DSF 50 \mu M) = 0.0007). h Survival curve of mice bearing GIC19-derived (left) and GIC31-derived (right) intracerebral xenografts treated with vehicle (GIC19: n = 8, GIC31: n = 5) or DSF (GIC19: n = 10, GIC31: n = 6). Log-rank (Mantel–Cox) test, two-sided, p value (GIC 19) = 0.0450; p value (GIC 31) = 0.8352. All graphs report mean ± SEM. Statistical significance for all panels *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001. Source data are provided in the source data file.
of its potential clinical value. Importantly, the epigenetic regulation of PTGER4 gene expression by DNA methylation could represent a valuable biomarker to be further developed as a predictor of drug response, particularly taking into account the roll out of methylation-based classification approaches in brain tumours.

More work will be required to assess whether the SYNGN (syngeneic comparison of GIC and iNSC) pipeline can be used in the clinical management of patients with GBM in a timeframe suitable for adjuvant treatment and whether syngeneic GLICO and intestinal organoids accurately predict tumour response and non-neoplastic cell toxicity on a patient-specific basis (Fig. 8). It

Fig. 7 Comparative analysis of transcriptome and methylome of GIC/iNSC identifies NTRK2/CTX as patient-specific drug sensitivity. a Drug treatment of GIC (black curves) and iNSC (grey curves) of patient 30 (top) and 18 (bottom) with doses ranging from 1 nM to 10 µM of CTXB (blue dots) or vehicle (black dots). Results are expressed in percentage of viability on the vehicle and were measured at end point, area under the curve (AUC) was calculated from percentages of viability (n = 6 experiment repetitions, two-tailed t-test, p value (patient 30) = 0.0326, p value (patient 18) = 0.5394). b Proliferation assay of GIC30 (top) and 18 (bottom) treated with vehicle (black bars) or CTX-B (blue hatched bars) assessed by cell numbers after 4 days of culture. Results are expressed as fold change of cell number at day 4 on day 1 of treatment and standardized on the vehicle treatment conditions (n = 3 experiment repetitions, two-tailed t-test, p value (GIC 30) = 0.0080, p value (GIC 18) = 0.9363). Viability of total cells of SYNGLICO (cerebral organoids (CO) + GICGFP) after 4 days of treatment with vehicle or CTX-B (50 µM, 100 µM) (c) and GFP Mean Fluorescence Intensity (MFI) of GICGFP+ cells (d) (COGIC30 vehicle: n = 10, CTX-B 50 and 100 µM: n = 9, COGIC18: n = 7 SYNGLICOs per group, one way ANOVA, two-sided. c Patient 30: p value (Vehicle vs CTX-B 100 µM = 0.3759); d Patient 30: p value (Vehicle vs CTX-B 100 µM = 0.0230). e Survival curve of mice bearing GIC30-derived (left) and GIC18-derived (right) intracerebral xenograft treated with vehicle (n = 7) or tat-Cyclotraxin-B (n = 8), log-rank (Mantel–Cox), two-sided, p value (GIC 30) = 0.0290. All graphs report mean ± SEM. Statistical significance for all panels *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001. Source data are provided in the source data file.
For iNSC induction, a commercially available kit was used following the manufacturer’s protocol (Gibco, #A1647801). Briefly, EPSC colonies were harvested, feeders’ removal was performed and EPSC plated at density from 0.25 to 0.75 x 10^6 per well of six well geltex coated plates in EPSCM and Rock inhibitor 10 μM (Stem cell technology #Y27632). The next day, the media was replaced with Gibco neural induction media (Neurobasal media #211050, pen-strep 1X and neural induction supplement 1X #A1647801, rock inhibitor 10 μM), then changed every 2 days until day 7, cells were passaged at the density of 1 x 10^5/cm^2 in Gibco neural expansion media (Neurobasal 0.5X, Advanced DMEM/F12 0.5X #12634, pen-strep 1X and neural induction supplement 1X, rock inhibitor 10 μM), then media was changed every 2 days, cells were dissociated with accutase (Millipore #50010025) and replated 2–3 times to establish the iNSC lines. Cells were frozen in synth-a-Freeze cryopreservation medium (Gibco #A12542), GIBCO® Human Neural Stem Cells (H9 hESC-Derived, #N7800-100) were used as control, in StemPro® NSC SFM (Cat. no. A10509-01).

For differentiation of iNSC along the astrocytic and neuronal lineages, iNSC cells were plated on geltex (Gibco #A143300) coated culture dishes. After 2 days, the media was changed to neural differentiation media (Neurobasal media #21103, B27 serum-free supplement 1% Gibco #17504) and changed every other day until day 10. iNSC differentiation along the oligodendrocytic lineage was performed following the protocol developed by ^2 with some adaptations. Briefly, at day 7, neural expansion media was changed to N2 media made of 47.5 ml of Advanced DMEM/F12 (Invitrogen, #11320033), 0.5 ml of non-essential amino acids (Invitrogen, #11140035), and GlutMax (Invitrogen, 35050038), 50 µl of β-mercaptoethanol, 0.5 ml of N2 supplement (Invitrogen, #17502048) changed daily with fresh 100 mM Retinoic acid (RA) (Sigma, #R2625) and−: 1 µM Smoothened Agonist (SAG) (Millipore, #566660) until day 4, then replaced by N2B27 media supplemented with 1 ml of B27 supplement (Invitrogen, #A15287010) and 114 µl of insulin (Sigma, #I9278) changed daily with fresh RA and SAG. Cells were then scraped, and the single cells suspension was transferred to an ultra-low attachment 24-well plate (Corning Cat. #3473) in N2B27 media and changed every other day for 10 days. On day 22, cell aggregates were plated onto Poly-L-Ornithine and Laminin coated plates and cultured in adherent conditions for 45 days with PDGF media (Millipore, #566660) changed daily with 5 μg/ml insulin and 5 μg/ml EGF, and changed every other day.

Dissociated cells were plated on laminin-coated 6-well plate in NeuroCult NS-Abroblast media (STEMCELL, #07571), heparin (2 µg/ml; Gibco 1258-0710), mEGF (20 ng/ml, Prepro Tech, 315-09) and hFGF (10 ng/ml; Prepro tech, AF-100-18B). Established cells were passaged when 70% confluent, frozen in Stem Cell Banker (Ambion ZENOAQ, 11890), and stored in liquid nitrogen. Cell lines used in this study are primary lines derived from human tumours, they have been characterized by transcriptomic profiling and cultured according to current practice, including contamination assessment.

Fibroblasts were established from small strips of dura mater, collected at the surgery. The tissue was minced with a scalpel then spun down and resuspended in trypsin for 5 min at 37°C. To stop the reaction fresh fibroblast media (DMEM, 4.5 µM glucose, fetal calf serum, 2% L-Glutamine, and 1% Penicillin-streptomycin) was added. Samples were centrifuged and resuspended in fresh media, then plated in six well plates (Corning #B1201). Media was topped up frequently during the first week then changed every other day.

Fibroblasts reprogramming was performed following a published protocol^21. Briefly, fibroblasts were electroporated with reprogramming vectors expressing Oct4, c-Myc, Klf4, Sox2 (OCT3, 05 pl, 5 µg) and Rarg, Lfh1 (RL 2F, 5.0 µg) using Amaxa Nucleofector (Lonza, Germany) then plated on SNL feeders with M15 media (Knockout DMEM Invitrogen, 15% Fetal Bovine Serum HyClone, 1X Glutamin-Penicillin-Streptomycin Invitrogen, 0X non-essential amino acids Invitrogen). When colonies appeared, media was replaced with EPSCM (DMEM/ F12 Invitrogen, 20% Knockout Serum Replacement Invitrogen, 1X Glutamin-Penicillin-Streptomycin, 0X non-essential amino acids Invitrogen, 0.1 mM β Mercapto-ethanol Sigma, 106 U/ml hLIF Millipore supplemented with the following inhibitors: CH39921 Trycys 1 µM, JNK Inhibitor VIII Tocris 4 µM, SB203580 Tocris 10 µM, A-459259 Santa Cruz 1 µM and XAV939 Stratgene 1 µM). Colonies were picked and plated in 24 well SNL feeders plates for expansion and characterization.

For embryoid bodies (EB) generation, EPSC lines were harvested and feeders removal was performed in 20% foetal serum for 30 min at 37°C to allow feeders to attach to the bottom of the dish. Floating EPSC harvested were transferred in ultra-low attachment 96 well plates at different densities (60, 45, 35, and 20 K) in EB media (DMEM/ Knockout Invitrogen #108225-018, Fetal Bovine serum Gibco #16141061, nonessential amino acid Invitrogen #11140, glut-pen-strep Invitrogen #10378 and β mercaptoethanol) inactivated every other day. After 7 days, they were transferred in 24 well plates on gelatin coated glass coverslips and after 10–14 days, cells were fixed with PFA 4%, 30 min at room temperature, and immunocytochemistry was performed.
Co-culture with patient-derived GIC-labelled GIC. GIC were transduced with pLooGFP (Addgene, #12241) as below (see lentiviral production). On day 1 of co-culture, immortalized xenograft organoids generated from the patient-specific EPSC were transferred into each well of a 96-well round-bottom ultra-low attachment plate. The excess medium was removed and 50,000 GIC expressing GFP were plated in each organoid-containing well in GIC medium. On day 2 of co-culture, the excess medium was removed from each organoid-containing well and an additional 50,000 GIC expressing GFP were added. GIC media, and plates were incubated overnight. On day 3, each tumour-bearing organoid was transferred into a well of a 12 well super-low attachment NuncIon Sphere plate (Thermo Scientific) and maintained in cerebral organoid maturation media on an orbital shaker at 130 rpm for 5–7 days followed by 4-day drug treatment.

Generation and characterization of EPSC-derived intestinal organoids. Intestinal organoids were generated and cultured as per the manufacturer’s instructions (StemCell Technologies, #05140). Briefly, feeder-free EPSC were cultured in matrigel coated 24 well plates and once they reached confluence, endoderm differentiation was induced for 3 days, followed by 3–7 days of definitive endoderm differentiation until formation and release of spheroids from the monolayer. Mid/hindgut were then embedded in a dome of matrigel in a new Nunclon Delta surface-treated 24 well plate with flat bottom and cultured in an intestinal organoid medium, changed every 3–4 days. Organoids were passaged once a week at the density of 50 organoids/matrigel dome.

Animal procedures and murine cells cultures. All procedures were performed in accordance with licences held under the UK Animals (Scientific Procedures) Act 1986 and later modifications and conforming to all relevant guidelines and regulations.

Cell cultures and assays. Primary endogenous NSC (eNSC) were isolated from the sub ventricular zone (SVZ) of 3 months old C57Black6/J mice following established new Nunclon Delta surface-treated 24 well plate with

NaturE COmmuNications | https://doi.org/10.1038/s41467-021-26297-6 | www.nature.com/naturecommunications

Sequencing. RNA and DNA were extracted from cell pellets using the RNA/DNA/Protein Purification Plus kit (NORGEN, #47700), following the manufactured protocol. Mice and human RNA samples were sequenced on HiSeq4000, at 75 PE after polyA selection, by Oxford Genomics Centre. Human DNA methylation was assessed on the Illumina Infinium MethylationEPIC kit at the UCL Genomics, Institute of Child Health, and mouse DNA methylation was performed by Reduced Representation Bisulfite Sequencing (RRBS). For RRBS library preparation DNA samples were digested overnight with restriction enzyme MspI (Agilent), and then prepared following the manufactured protocol of NEBNext Ultra II DNA Library Prep Kit (NEB) and bisulfite conversion following Zimo kit (Quagen, D5001) and run on NextSeq 500 Mid Output Run 150 cycles (Genome Centre, Barts and the London School of Medicine and Dentistry).

Tregs migration assay. PBMCs were isolated from the blood of healthy donors (REC 17/LO/1061). They were collected between the top plasma layer and the bottom ficoll medium of a Histopaque1077 (Sigma-Aldrich) gradient and resuspended in PBS, washed and plated in a petri dish and RPMI 1640 + 10% FBS incubated 45 min to deplete macrophages. The supernatant was then harvested and PBMC were counted and frozen in FBS 90% and DSMO 10%. For ChABC (Sigma) treatment, the compound was added to cell culture media at the final concentration of 0.01 µM for 48 h. Then cells were stained for chondroitin 56 (Abcam) or harvested to be plated in transwell chamber.

Conditioned-media from GICs or iNSCs or cells (20 K) were added to the gelrex coated top chamber of a transwell 5 mm pore (Sarstedt, 83.392.500) on a 24 wells plate. Twenty-four hours later, PBMCs (300 K cells) in RPMI with FBS 10%, were added to the top chamber for a 4 h incubation at 37 C, CO2 5%. Cells were harvested in the bottom chamber, counted, and stained by flow cytometry.

Analysis of GAG composition using AMAC-labelling and RP-HPLC. Cultured GICs and iNSCs were washed with PBS then lysed using 1% (vol/vol) Triton X-100 for 2 h at room temperature. Conditioned-media and cell extract samples were treated with 100 µg/mL Pronase (Roche Diagnostics) for 4 h at 37 °C. Cell extracts were incubated for 15 min at 90 °C to deactivate the Pronase, before treating with 14 µg/mL DNAse I (Sigma) and 10 mM MgCl2 overnight at 37 °C to prevent DNA interference with downstream processing. All preparations were loaded onto a 1.5 mL diethylaminoethyl (DEAE)-Sephacel column (Sigma) and washed with 50 mL of 0.25 M NaCl, 20 mM NaH2PO4.H2O (pH 7.0) to remove hyaluronan. The sulphated GAGs were eluted with 5 mL of 1.5 M NaCl, 20 mM NaH2PO4.H2O (pH 7.0) and desalted using PD-10 G-25 pre-packed columns (Amersham Healthcare). Purified GAGs were freeze-dried, then digested with GAG lysates specific for either HS or DS for 12 h at room temperature. HS chains were digested using 0.8 mU of each heparanase I, II, and III (from Flavobacterium heparinum, Ipurin) in 100 µL of 0.1 M sodium acetate, 0.1 mM calcium acetate (pH 7.0). CS/DS chains were digested using 0.8 mU of either chondroitinase ABC (chABC) (Sigma) in 100 µL of 50 mM Tris, 50 mM NaCl (pH 7.9). The resulting disaccharides were freeze-dried then re-suspended in 10 µL of 0.1 M AMAC and incubated at room temperature for 20 min. To each reaction, 10 µL of 1 M NaBH4CN was added and incubated at room temperature overnight in the dark. The AMAC-labelled samples were separated by RP-HPLC in duplicate using a Zorbax Eclipse XDB C18 RP-HPLC column (35 µM, 2.1 x 150 mm) (Agilent Technologies), following published protocols. Disaccharides were detected by fluorescence and were identified and quantified in comparison with known standards of commercial disaccharide standards (Ipurin).

Immunoﬂuorescence and immuno histochemistry. Staining EPSC, NSC, and EB in adherent cultures on a glass coverslip was washed once with PBS then fixed with 3% paraformaldehyde for 30 min at room temperature. After three washes with PBS, blocking with 3% Donkey normal serum, 0.3% Triton X100 PBS for 30 min at room temperature was performed prior to incubation with primary antibodies (see Table S3) at 4° overnight. After three washes with PBS and 1 h incubation with secondary antibodies (see Table S3) diluted in PBS, cells were washed again with PBS, and coverslip were mounted on SUPERFROST slides with mounting medium containing DAPI for nuclear counterstaining (Vectorshied H-1000). Microscope analysis was performed on IN Cell Analyzer 2200 Cell Imaging.
Cerebral and intestinal organoids were cultured in FFA (4%), overnight and 30 min at room temperature, respectively. After washes with PBS, organoids were incubated in 30% sucrose until they sunk, embedded in 7.5% gelatin (in 10% sucrose), and then frozen in cold isopentane. Cerebral organoids and GLICO were cryosectioned at 20–25 µm and intestinal organoids at 10 µm using the Leica cryostat. All sections were rehydrated, permeabilized for 1 h at RT (4% serum, 0.25% Triton-X in PBS), then incubated in normal serum (4% serum, 0.1% Triton-X in PBS) and the relevant primary antibody overnight at 4°C. Primary antibodies for Sox2, Nestin, and NeuN (cerebral organoids) and Villin, E-Cadherin, β-catenin, and CDX2 (intestinal organoids), as well as Ki67 and cleaved-caspase3, were used (Table S3). Secondary antibodies were incubated for 2 h at room temperature, and slides were mounted with mounting media containing DAPI.

**Quantification.** Chondroitin sulfate fluorescence staining was analyzed by CTCF (corrected total cell fluorescence): CTCF = Integrated Density – (Area of selected cell × Mean fluorescence of background readings) with Image J software.

Subcutaneous tumour injected in vivo with Tregs GFP-labeled slides were scanned with TISSUEFAX® and GFP positive cells and nuclei were automatically counted with Staarquest software.

Cryosections of the organoids were stained with GFP and Ki67 or GFP and cleaved-caspase-3, and nuclear counterstained with DAPI as described above. For intestinal organoids, a protocol of all organoids were captured at high magnification, nuclei and positive cells were then manually counted. For cerebral organoids, 4–6 fields per condition for each individual biological sample were captured as Z-stacks of approximately 10µm at 63x magnification using the Zeiss 880 Laser Scanning Confocal Microscope with Fast Airyscan and MultiPhoton. Z-stack images were then Airyscan processed and analyzed using the IMARIS software. Briefly, each image was converted to a .ims file, each channel (GFP, RFP, DAPI) was set to a threshold in order to use the spot function analysis and applied to all images. GFP spots represent GIC, GFP spots Ki67 or Cleaved-caspase-3, and DAPI spots as cells (nuclei). Spot counts and double-positive overlapping spots results were then exported and analyzed in GraphPad Prism 8.

**Flow Cytometry.** Cells were washed once with PBS then plated and pelleted in 96 wells plate in 200 µl PBS. Antibodies CD4 PE, CD8-AF700, CD25-PECy7, CD127-APC, CD31-FITC, CD14 PerCP Cy5.5, Foxp3, and L/D marker Cy5.5 and L/D marker Cy5.5 were diluted in FACS buffer and incubated at room temperature for 30 min. Cells were then washed, permeabilized, and fixed with Transcription Factor Staining Buffer Set (eBioscience 00-5523-00). Foxp3 antibody was diluted in the permeabilization buffer and incubated for 20 min at room temperature. Finally, cells were washed, resuspended in FACS Buffer and run on BD LSRII device, with a FACSDivya software Version 8. Data were transferred and analyzed using FlowJo software V10 (Tree Star, Oregon, USA).

The gating strategy employed was as follows—cells were first selected based on size using the FSC-A and SSC-A. Then live cells were selected for by gating on cells negative for the membrane dye-ethyl 0.66. Gating on CD4 and CD8 markers. Gating for Treg population relied on CD4* Foxp3* CD25* CD127* marker expression, monocytes relied on CD14* population, natural killer on CD56* population while other immune T cells were relied on CD4* Foxp3* CD25* and CD8* Foxp3.

**CO-GIC cultures.** Upon dissociation with the Neural Dissection kit and staining with the Zombie NTR™ Fixable Viability Kit, all samples were analyzed using the BD LSRII flow cytometer. After exclusion of cell debris (FACS plots SSC-A/FSC-A), and gating of single cells (SSC-W/FSC-A), the overall viability of the organoid was assessed after gating for Zombie NTR™ negative cells (live cells), shown on plots as a histogram of Comp-APC-Cy7−A. From the single cell gating, the overall GFP+ population was detected (SSC-A/Comp-FITC-A) including both live and dead GFP-expressing GIC, from which the MFI was measured as the geometric mean of Comp-FITC-A. Within the live cell gating, cells were gated for their expression from which the MFI was calculated as described before. For co-cultures with epidermally edited GIC18, carrier the mCherry staining was measured. Gating for mCherry staining, samples were analyzed and gated in the same manner to assess viability. In order to assess the MFI of GIC, double-positive cells for BFP and mCherry were gated, from which the MFI of BFP was analyzed.

All FACS analysis was done using FlowJo v10.6.2. All treated co-cultures’ viability and GFP+ normalization to the vehicle control of the respective condition and for each time point, the experiment was performed twice. All results of biological replicates were analyzed using OneWay-ANOVA in GraphPad Prism 8.

**Drug treatment.** In vitro drug treatments in 2D cultures were performed on 2k cells plated in 96 well plates; 4 days treatment was performed with a range of doses from 0.5 µM to 10 µM and 100 µM PGF-1OH, 50 µM and 100 µM CTX-B or 10 µM and 50 µM DSF. Each drug was added to the final volume on day 1, followed by drug top-up every day for the following 3 days, without discarding any of the contents in each well. Vehicle control co-cultured organoids were treated with DMSO. Organoids were then processed and analyzed with the BD LSRII cytometer as described above and fixed with 4% paraformaldehyde. For FACS analysis on SYNGLICO model: On day 4 post-treatment, all contents of each well including the co-cultured cerebral organoids were collected into tubes and the Neural Dissection kit (MACS Miltenyi Biotec) was used to dissociate organoids into single-cell suspension.

Drug treatment in xenograft mice model started two months after intra-cranial injection of GIC, 5 times a week in IP with DSF (100 mg/kg) or vehicle (DMSO 5%) and tur-Cyclostatin B (Bio S&T, 20 mg/kg) of vehicle (saline solution) (Cazorla et al., 2010). Mice were weighed 3 times a week to adjust doses.

**Plasmid design and construction.** FUW-dCas9-Tet1-CD construct plasmids and the pgRNA (empty scaffold plasmid guide RNA) were a gift from Dr. Rudolf Jaenicke. All PGF4 gRNA plasmids were designed based on the sequence near the hypermethylated CpG islands identified in syngeneic comparison (Fig. 4b). (CpG1: cg10603323, CpG2: cg24716010, CpG3: cg11408333, CpG4: cg46013125, CpG5: cg03332743, CpG6: cg17421097, CpG7: cg20062698, CpG8: cg01822716, CpG9: cg21149775, CpG10: cg10952088, CpG11: cg01987756, CpG12: cg01426707). Six guides were designed using the Ensembl GRC37 reference. In order to close the target sequences into the pgRNA, two complement oligonucleotides were designed (see Table S3). To ensure cloning compatibility, TTGG- and AAC- were added to the oligonucleotides sequence (5'-3'). Oligos were designed to be cloned in the guide RNA plasmid to specifically fused proteins in a long-overlapping manner.

Oligonucleotides duplexes were formed to establish single guide RNA (sgRNA), by oligonucleotide phosphorylation using the T4 Polynucleotide Kinase (PNK, ThermoFisher). 100 µM of each gRNA sequence (forward and reverse) with 25 mM ATP and PNK/Reaction buffer (10x) were added in a PCR tube, followed by 30 min incubation at 37°C, 5 min at 95°C, and ramp to 4°C by 0.1/s to join the single 5′ at the 3′ end of the reverse oligo annealed the sequence to the additional G added to the forward oligo. The pgRNA vector was cut and modified with the Aarl enzyme (ThermoFisher) targeting the Aar1 sites (CAACCCG) at 37°C incubation overnight. The enzyme was then inactivated at 65°C for 20 min. The gRNA plasmids were cloned by inserting annealed oligonucleotides into Aar1 site modified pgRNA plasmid. After ligation of the annealed duplexes into the digested vector, the cut plasmid was added along with the duplex oligos, ligation buffer, and quick ligase (Quick Ligation kit, NewEngland Biolabs) for 5 min at 25°C. To ensure only circularized DNA was present, the ligation product was digested with plasmid-safe ATnPDependent DNAses at 37°C for 30 min to digest any unincorporated DNA.

**Lentiviral production.** For the production of lentiviruses expressing the two FUW plasmids dCas9-Tet1-CD (catalytic domain), and all gRNA plasmids, pLOXGP plasmid (Addgene), plasmid for overexpression of PTGER4 (Horizon-Dharmacon) and shNTRK2 plasmid (Horizon-Dharmacon), 5 × 10^4/cm^2 HEK 293T cells were transfected with lipofectamine 3000 (Invitrogen, L3000), the lentiviral plasmids pCMV-G, pCMV-HIV1 (Table S3) and the plasmid of interest. After 48 h, the supernatant was collected, cell debris was removed by filtration (0.45 PVDF filter, Sartorius) and the lentiviral particles were precipitated with polyethylene glycol (Sigma Aldrich) and stored at 4°C for 16 h. Lentiviral particles were concentrated by centrifugation (30 min, 15000 x g, 4°C), resuspended in sterile PBS, and stored at −80°C. Transducing units were then determined by plating 0.5 × 10^5 HEK 293T cells into 12-well dishes and exposing them to serial dilutions 10^−1 to 10^−5 of the lentiviral supernatant. Cells were collected 96 h from the transduction and the percentage of GFP-positive cells was determined by FACS BD FACs Canto II). The titration in transducing units per ml (TU/ml) was calculated according to the following formula: [(Positive cells/100) n x transduction volume of virus (ml)] / (100 x transduction volume of virus (ml)) = (Positive cell count) / (100 x transduction volume of virus (ml)) = (TU/ml) x 100. For proliferation assay in drug treatment, PGE1-OH (10 µM), Cyclostatin-B (10 µM), DSF (1 µM), or vehicle were added to the media the day after the transduction.
were seeded. Media and drugs or vehicle were topped up every day and cells were harvested and counted at days 3, 7, and 10 and manually counted.

**Apopotosis.** Cells were plated (30 K) in 24 wells plates and treated 4 days with vehicle or drug (DSF 1 µM, Cyclotransit-B 10 µM or PGE1-OH 10 µM). Apoptosis was measured at end point following the manufacturer’s protocol (Caspase-3 assay kit,Abcam, ab39401). Briefly, cells were detached and resuspended in Cell Lysis Buffer for 10 min on ice, centrifuged and protein concentration was assessed from the supernatant. DTT was mixed with the reaction Buffer and after 10 min incubation at 37°C, the sample was centrifuged and washed with PBS in ice and resuspended in Annexin buffer. The cells were then incubated with Annexin V-fluorescein isothiocyanate (FITC) and propidium iodide (PI) (1:1) for 15 min at room temperature in the dark. Then, the cell suspension was assayed by flow cytometry using FACSCanto II.

**Sphere-forming assay and limiting dilution assay.** Cells were plated (30 K) in 24 wells plates with vehicle or drug (DSF 1 µM, Cyclotransit-B 10 µM or PGE1-OH 10 µM) in triplicate and cultured for 10 days as sphere-forming assay and limiting dilution assay. Cells were washed, detached with accutase, and 1 × 106 cells were resuspended in standard error (SEM).

**RT qPCR.** Protein expression was performed by densitometric analysis and normalized with PGE1-OH 10 µM, then washed and detached with accutase and seeded (1 × 104) in 24 wells plates with vehicle or drug (DSF, CTX-B, and PGE1-OH) were added every day for 10 days on GIC line of interest and one control line for each drug. Each sample was run in triplicate. 

**Western blot.** Protein homogenates were obtained using RIPA buffer containing protease cocktail inhibitors (Santa Cruz Biotechnology). Protein concentration was determined by BCA assay (Pierce) or two-way ANOVA as appropriate, and displayed as the mean ± standard deviation (SD).

**RNA-Seq.** Reference samples were downloaded from OmicsBase GEO repository (GSE75300: Kang’s NSC; GSE43916: Wapinski’s NSC; GSE119471: Zhang’s NSC; GSE54141: Friedmann-Morvinski’s astrocytes and neurons; GSE57992: Schmid’s fetal NSC; GSE89898: Yanez’s monocytic lineage see the main text; for references see main text) and were processed together with the samples from this study (6 eNSC, 9 iNSC Gibco and 9 iNSC N2B27): quality was assessed via FastQC and trimmng of low-quality reads and adapters was performed via TrimGalore v. 0.4.5-1 with default parameters. The pseudosignal packaging Salmon v. 0.13.1 was used to map all samples to a previously indexed reference genome Ensembl GRCh38 (release 90) to obtain transcript per million (TPM) expression levels, normalized for library size and transcript length. Transcript to gene quantification was performed via bioMAq in R and low expressed genes with TPM < 1 were filtered out. The expression values were finally z-transformed prior to plotting.

**Statistical analysis for the wet lab experiments.** Sample processing was carried out blinded. Statistical analysis was performed using GraphPad software unless otherwise stated. Significance was determined with t-test, one-way ANOVA (with Sidak’s multiple comparison test) or two-way ANOVA as appropriate and displayed as mean ± standard error (SEM). p < 0.05 was considered significant. Significance was indicated with asterisks: *p < 0.05; **p < 0.01; ***p < 0.001. All variables were assumed to be normally distributed unless otherwise stated. Outliers were considered those data points furthest from the median value.

**Computational analysis (human samples).** RNA-Seq. RNA-Seq data were processed using two separate pipelines. For hierarchical clustering and principal component analysis, transcript-level expression was estimated directly using the pseudocount package Salmon78. The results were then normalized to obtain gene-level expression estimates in units of transcripts per million (TPM), which are normalized for library size and transcript length. For the purpose of identifying deregulated genes, gene counts were first estimated using the gapped alignment software STAR2. The reference genome used in both pipelines is Ensembl GRCh38 (release 90). DE genes are computed using the package edgeR80 with the command edgeR. Baseline dispersion is first estimated across all samples, before applying a two-sided quasi-likelihood F test to estimate gene-wise differential expression separately for
each patient, following the suggested best practices of the authors. P values are corrected for multiple hypothesis testing using the Benjamini-Hochberg correction, resulting in an expected false discovery rate (FDR) of < 0.01.

**DNA methylation.** All DNA methylation data generated as part of this study were profiled using the Illumina HumanMethylation EPIC array. In addition, we included reference lines profiled on both the EPIC and Illumina HumanMethylation 450 K arrays. When comparing the two, EPIC array data were reduced to include only the same probes as the 450 K data (these have the same chemical design and are therefore comparable). Raw array data were first pre-processed using the CHAMP package in R to remove failed detections and probes with known design flaws, before normalization using the SWAN algorithm. Differentially methylated regions are inferred following the strategy in the R package DMRForPairs. Briefly, clusters are defined as groups of six or more probes separated by no more than 400 base pairs. Each cluster is then assessed for biological significance by requiring an absolute median change in the M value (ΔM) greater than or equal to 1.4 between two comparison groups (e.g., iPSC vs ESC).

Where replicates are present in either comparison group, we take the median value for each probe across the samples in the group before computing ΔM. Where the cluster is deemed biologically significant, we assess statistical significance using the Mann–Whitney U test, comparing all probes in each condition. P values are adjusted for multiple hypothesis testing using the Benjamini–Hochberg FDR and the cluster is declared a DMR if it has an FDR < 0.01.

The demethylation analysis for PTGER4 via grNA was carried out in a similar way: data were profiled using the Illumina HumanMethylation EPIC array, while the R CHAMP package was used to remove flawed probes and failed detections. Finally, the SWAN algorithm was used for normalization before comparing the results. **Classifying EPR lines:** The success of the reprogramming process was assessed using gene expression and DNA methylation data as follows. For DNA methylation, we compare our EPs lines and reference iPSC lines with two reference ESC lines (see main text for details of all references). For each ESC/iPSC line, we extracted core DMRs that were present in both ESC comparisons (i.e., the intersection of the two DMR lists). These were used to create the pair plots in Figs. S2e and 2f. For gene expression, we took an analogous approach, defining core DE genes as those present in both ESC comparisons. These were used to create the plot in Fig. 5h.

Computing GBM subgroup based on DNA methylation profile: Subgroups of GBM have previously been described based on analysis of methylome data. Our bulk FFPE and GIC samples were assigned to subgroups as defined above on this basis of their methylation profiles using a published random forest classifier. This accepts raw methylation array data as an input. Integrative analysis of RNA-Seq and DNA methylation: DE and DMR results were compared to the annotation for the Illumina HumanMethylation EPIC array used to obtain the DNA methylation results. In this annotation, a list of corresponding genes is provided for each CpG probe on this array together with the way in which the CpG probe is related to that gene (e.g., “falls within gene 18 exon”). A CpG probe is considered relevant to a gene if it is anywhere within the unspliced gene (i.e., including intronic regions) or within 1.5 kb of the transcription start site. Each DMR comprises 6 or more CpG probes. In order to integrate DE and DMR results, we pooled all relevant genes for these probes. A DE/DMR pair is considered concordant if the effect is in opposite directions (hypermethylated and hypomethylated) and upregulated/downregulated.

We elected to focus on patient-specific gene targets identified as differentially expressed and methylated in just one patient GIC line relative to syngeneic iNSC. In order to prioritize this large list of genes, we imposed increasingly stringent requirements, defined in the main text. The lower tables in Fig. 4a show toy example for the four patients used. Bold outlines indicate statistical significance. Gene A is not patient-specific at the level of DE or DM. Gene B has patient-specific DE but no corresponding DM. Gene C has patient-specific DE and is DM in multiple patients; this gene is the longest. Gene D has patient-specific DE and concordant patient-specific DM; this gene is the shortlist.

**Computing enriched pathways in DE gene sets.** The Ingenuity Pathway Analysis (IPA) platform was used to compute enriched gene sets in DE gene lists. All pathways with an uncorrected p value < 0.05 were retained, though stricter thresholds were applied in subsequent analyses. The genes involved in enriched pathways were exported separately from IPA in order to generate network visualizations and to check for cross-talk between IPA signatures and the cell type signatures used in xCell.

In order to compare the extent of pathway enrichment between syngeneic DE gene lists and DE gene lists arising from a comparison of our GIC with a reference NSC line (Fig. S5b), we retained all pathways with p < 0.005. For each pathway, we computed the number of patients with a DE list enriched for that pathway (GIC relative to syngeneic iNSC, H9 NSC or Gibco NSC). The resulting vectors were compared using the Wilcoxon signed-rank test in order to determine whether the number of pathways captured with each comparator differed significantly. The rank correlation metric was used to determine DE lists captured more pathways (i.e., the effect direction). We found that the syngeneic DE lists captured significantly more pathways than using the H9 (p = 0.029) or Gibco (p = 0.003) reference comparators. A similar but more significant difference is obtained if the sum of −log10(p) values across patients is used instead of the binary yes/no metric described here.

Network analysis: CytoVista version 3.7.20 was used to perform network analysis on the significantly enriched pathways (uncorrected p value < 0.01) found with IPA (QIAGEN), in the 10 syngeneic GIC-iNSC transcriptomic comparisons. Enrichment map and Autoannotate were used for visualization and clustering, respectively. The edge similarity threshold was set at 0.5. Various layout modifications were implemented, such as edge bundling.

Inferring cell-type proportions in bulk tissue: Gene expression quantification (10x v2.0) of bulk GBM FFPE samples was processed using the xCell web application in order to infer the proportions of 64 immune and stroma cell types within each bulk sample. This inference is based on a curated set of one or more gene signatures corresponding to each cell type, ranging 489 in total. For the purpose of eliminating cross-talk between IPA pathway signatures and xCell, we also downloaded these signatures.

Correlating cell type proportions with pathway enrichment: Before analyzing the correlation between inferred cell-type proportions and IPA pathway enrichment results, it was first necessary to exclude combinations for which the cell type signature overlaps extensively with the pathway. Therefore, any occurrence of an IPA pathway for which >10% of the genes were also included within a cell type signature was ignored. For each cell type reported by xCell and pathway reported by IPA, we computed the Spearman rank correlation and associated exact p value across all patients between the −log10(p) values from the pathway enrichment and the inferred cell type proportion. As this correlation is rank-based, there is no need to correct pathway enrichment p values for multiple hypothesis testing. An example of this is such comparisons is shown in Fig. S7a.

DNA methylation and RNA expression data from the independent GIC cohort HGCC, publicly available, were used to assess the proportion of patients with PTGER4, NTRK2, and ALDH3B1 signatures in a wider population of patients. For the differential gene expression analysis of GICs from the HGCC cohort, microarray data that had been pre-processed (normalized and Combat batch adjusted) was provided by our collaborator Professor Sven Nelander, Uppsala University. Differential gene expression analysis was then performed in an R workspace using the Rstudio environment and limma package 83 (version 3.40.6). DNA methylation was processed with Illumina array, as described above, and values were used for further analysis. Briefly, for each of the three genes, average RNA sequencing and DNA methylation value were calculated and patients with high methylation and low expression for PTGER4 were identified while patients with low methylation and high expression were selected for NTRK2 and ALDH3B1.

Schematics were created with BioRender.com

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

**Data availability.** The authors declare that all data supporting the findings of this study are available within the article, its supplementary information files, and the public repositories as stated below. The datasets generated in this study are available as raw data in the NCBI Gene Expression Omnibus database with SuperSeries number GSE15994, which includes all mouse data (GSE154367, the human transcriptomic GSE154958, and the human microarray data GSE155995). Source data are provided with this paper.

PUBLICLY AVAILABLE DATASETS USED IN THE STUDY: GSE124532, GSE1870, GSE78690, GSE60054, GSE6280, GSE75300, GSE43916, GSE119741, GSE64411, GSE75592, https://www.bionformatics.babraham.ac.uk/. Source data are provided with this paper.

Received: 2 March 2021; Accepted: 23 September 2021; Published online: 21 October 2021

**References.**

Stupp, R. et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* **10**, 459–466 (2009).

Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).

Cancer Genome Atlas Research N. et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

Sturm, D. et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).

Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56 e46 (2017).
6. Gimple, R. C., Bhargava, S., Dixit, D. & Rich, J. N. Glioblastoma stem cells: lessons from the tumor hierarchy in a lethal cancer. *Genes Dev.* 33, 591–609 (2019).

7. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15, 81–94 (2018).

8. Singh, S. K. et al. Identification of human brain tumour initiating cells. *Nature* 432, 396–401 (2004).

9. Costanzo, C. P. et al. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* 11, 3406 (2020).

10. Alcantara Llaguno, S. et al. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer Cell* 15, 45–56 (2009).

11. Holland, E. C. et al. Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice. *Nat. Genet.* 25, 55–57 (2000).

12. Lee, J. H. et al. Human glioblastoma arises from subventricular zone cells with low-level driver mutations. *Nature* 560, 243–247 (2018).

13. Lan, X. et al. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* 549, 227–232 (2017).

14. Nefef, C. et al. An integrative model of cellular plasticity, and genetics for glioblastoma. *Cell* 178, 835–849 e821 (2019).

15. Mack, S. C. et al. Chromatin landscapes reveal developmentally encoded transcriptional states that define human glioblastomas. *J. Exp. Med.* 216, 1071–1090 (2019).

16. Zhou, D. et al. D2F04: a distinctive epigenetic signature in glioblastoma stem cells and their response to differentiation cues. *Genome Biol.* 19, 43 (2018).

17. Hegi, M. E. et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *Nat. Med.* 10, 997–1003 (2004).

18. Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* 455, 469–474 (2008).

19. Verhaak, R. G. W. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cell Cancer J.* 17, 98–110 (2010).

20. Weinberg, E. et al. An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep.* 3, 1567–1579 (2013).

21. Pollard, S. M. et al. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* 4, 568–580 (2009).

22. Dumas, A. A. et al. Microglia promote glioblastoma via mTOR-mediated immunosuppression of the tumour microenvironment. *EMBO J.* 39, e103790 (2020).

23. Yang, J. et al. Establishment of mouse expanded potential stem cells. *Nature* 550, 393–397 (2017).

24. Chin, M. H. et al. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5, 111–123 (2009).

25. Miletic, H. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, 68–73 (2011).

26. Ruiz, S. et al. Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* 109, 16196–16201 (2012).

27. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* 28, 1045–1048 (2010).

28. Cortes, J. et al. Human pluripotent reprogramming with CRISPR activators. *Nat. Commun.* 9, 2643 (2018).

29. Banovich, N. E. et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res.* 28, 122–131 (2018).

30. Caccia, R. et al. Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. *Cell* 162, 412–424 (2015).

31. Koyt, I. et al. High-efficiency RNA-based reprogramming of human primary fibroblasts. *Nat. Commun.* 9, 745 (2018).

32. Choi, H. et al. In vivo differentiation of induced pluripotent stem cells into neural stem cells by chimera formation. *PLoS One* 12, e0170735 (2017).

33. Lapinska, O. L. et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* 155, 621–635 (2013).

34. Zhang, Y. et al. Mutant IDH sensitizes gliomas to endoplasmic reticulum stress and triggers apoptosis by MicroRNA183-mediated inhibition of Semaphorin 3E. *Cancer Res.* 79, 4994–5007 (2019).

35. Kang, Y. J. et al. Core pathway mutations induce de-differentiation of murine astrocytes into glioblastoma stem cells that are sensitive to radiation but resistant to temozolomide. *Neuro Oncol.* 18, 962–973 (2016).
84. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: subset-quantile within array
85. Rijlaarsdam, M. A., van der Zwan, Y. G., Dorssers, L. C. & Looijenga, L. H.
83. Feber, A. et al. Using high-density DNA methylation arrays to pro
86. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical
2020.
87. Shannon, P. et al. Cytoscape: a software environment for integrated models of
68. Soares da Costa, D., Reis, R. L. & Pashkuleva, I. Sulfation of
glycosaminoglycans and its implications in human health and disorders. Annu
88. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for
69. Logun, M. T. et al. Surfen-mediated blockade of extratumoral chondroitin
sulfate glycosaminoglycans inhibits glioblastoma invasion. FASEB J. 33,
11973–11992 (2019).
89. Weyers, A. et al. A structural analysis of glycosaminoglycans from lethal and
nonlethal mice: does the presence of high heparan sulfates explain the difference in
86. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical
2018.
82. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner.
83. Feber, A. et al. Using high-density DNA methylation arrays to pro
78. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor
1992 (2019).
87. Shannon, P. et al. Cytoscape: a software environment for integrated models of
1147–1148 (2014).
88. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for
79. Chen, Y., Pal, B., Visvader, J. E. & Smyth, G. K. Differential methylation
analysis of reduced representation bisulfite sequencing experiments using edgeR.
F1000Res 6, 2055 (2017).
80. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for
ChIP peak annotation, comparison, and visualization. Bioinformatics 31,
2382–2383 (2015).
81. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon
provides fast and bias-aware quantification of transcript expression. Nat.
Methods 14, 417–419 (2017).
82. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29,
15–21 (2013).
83. Feber, A. et al. Using high-density DNA methylation arrays to profile copy
count alterations. Genome Biol. 15, R30 (2014).
84. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: subset-quantile within array
normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome
Biol. 13, R44 (2012).
85. Rijkersdam, M. A., van der Zwan, Y. G., Dorsers, L. C. & Loosjenga, L. H.
DMRforPairs: identifying differentially methylated regions between unique
samples using array based methylation profiles. BMC Bioinform. 15, 141
(2014).
86. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical
and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B, 57, 289–300 (1995).
87. Shannon, P. et al. CytoScape: a software environment for integrated models of
biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
88. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for
ChIP peak annotation, comparison, and visualization. Bioinformatics 31,
2382–2383 (2015).
89. Weyers, A. et al. A structural analysis of glycosaminoglycans from lethal and
nonlethal mice: does the presence of high heparan sulfates explain the difference in
86. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical
and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B, 57, 289–300 (1995).
87. Shannon, P. et al. CytoScape: a software environment for integrated models of
biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
88. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for
ChIP peak annotation, comparison, and visualization. Bioinformatics 31,
2382–2383 (2015).
90. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon
provides fast and bias-aware quantification of transcript expression. Nat.
Methods 14, 417–419 (2017).
91. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29,
15–21 (2013).
92. Feber, A. et al. Using high-density DNA methylation arrays to profile copy
count alterations. Genome Biol. 15, R30 (2014).
93. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: subset-quantile within array
normalization for Illumina Infinium HumanMethylation450 BeadChips. Genome
Biol. 13, R44 (2012).
85. Rijkersdam, M. A., van der Zwan, Y. G., Dorsers, L. C. & Loosjenga, L. H.
DMRforPairs: identifying differentially methylated regions between unique
samples using array based methylation profiles. BMC Bioinform. 15, 141
(2014).
86. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical
and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B, 57, 289–300 (1995).
87. Shannon, P. et al. CytoScape: a software environment for integrated models of
biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
Acknowledgements
This work is funded by grants from Brain Tumour Research (Centre of Excellence award
to S.M.), Cancer Research UK (C23965/A29199 programme award to S.M.), Barts
Charity (MGU0447 programme grant to S.M.), The Brain Tumour Charity (GN-000389
clinical research training fellowship to T.M.), NIHR (CL-2019-19-001 to TM) and The
Willoughby Fund Trustees (studenthip to AAD). Part of the study was funded by the
National Institute for Health Research to UCLH Biomedical research centre (BRC399/
NS/RB101410 to S.B.). S.B. is also supported by the Department of Health’s NIHR
Biomedical Research Centre’s funding scheme. J.L.T. and C.L.R.M. are supported by the
Engineering and Physical Sciences Research Council (EP/N008615/1 programme grant)
and F.M.B by The British Heart Foundation (BHF RG/14/2/30616). We thank the staff
of our animal facility for help with the daily care of our mouse colony. We thank Dan
Pennington for advice on ethical requirements for blood collection for research use
and Reiss Browning for performing tail vein injections. We acknowledge the use of data
generated by the TCGA Research Network: https://www.cancer.gov/tcga.
Author contributions
C.V., L.G., J.R.B., M.C., A.A.D. and X.Z. designed and performed experiments and
analyzed results. G.R. and N.P. designed and performed all computational analyses. J.R.
T.E. and N.A. identified and consented the patients for the study. S.B. obtained ethical
approval and supervised human sample collection and pathological and molecular
analysis of the biopsies. Y.M.L. performed image analysis. T.O.M. assisted with
intracranial-injection experiments J.V., S.N.A. and F.M.B co-designed and supervised the
Tregs experiments. J.L.T. and C.L.R.M. performed the GAG experiments. T.J. assisted
with karyotyping. V.R., S.N., D.S., Y.Y.L. and P.L. shared datasets and essential expertise.
S.M. conceived and supervised the study. C.V., G.R. and S.M. wrote the manuscript with
contribution from all authors.
Competing interests
The authors declare no competing interests.
Additional information
Supplementary information The online version contains supplementary material available
at https://doi.org/10.1038/s41467-021-26297-6.
Correspondence and requests for materials should be addressed to Silvia Marino.
Peer review information Nature Communications thanks Monika Hegi and Ulf Kahler
for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permission information is available at http://www.nature.com/reprints
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in
published maps and institutional affiliations.
Open Access This article is licensed under a Creative Commons
Attribution 4.0 International License, which permits use, sharing,
adaptation, distribution and reproduction in any medium or format, as long as you give
appropriate credit to the original author(s) and the source, provide a link to the Creative
Commons license, and indicate if changes were made. The images or other third party
material in this article are included in the article’s Creative Commons license, unless
indicated otherwise in a credit line to the material. If material is not included in the
article’s Creative Commons license and your intended use is not permitted by statutory
regulation or exceeds the permitted use, you will need to obtain permission directly from
the copyright holder. To view a copy of this license, visit http://creativecommons.org/
licenses/by/4.0/.
© The Author(s) 2021