

RESEARCH

Open Access



# Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance

Christine E. Eyler<sup>1,2†</sup>, Hironori Matsunaga<sup>2,3†</sup>, Volker Hovestadt<sup>2,3</sup>, Samantha J. Vantine<sup>2,3</sup>, Peter van Galen<sup>2,3</sup> and Bradley E. Bernstein<sup>2,3\*</sup>

\* Correspondence: [bernstein.bradley@mgh.harvard.edu](mailto:bernstein.bradley@mgh.harvard.edu)

<sup>†</sup>Christine E. Eyler and Hironori Matsunaga contributed equally to this work.

<sup>2</sup>Broad Institute of Harvard and MIT, Cambridge, MA, USA

<sup>3</sup>Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Full list of author information is available at the end of the article

## Abstract

**Background:** Tumors can evolve and adapt to therapeutic pressure by acquiring genetic and epigenetic alterations that may be transient or stable. A precise understanding of how such events contribute to intratumoral heterogeneity, dynamic subpopulations, and overall tumor fitness will require experimental approaches to prospectively label, track, and characterize resistant or otherwise adaptive populations at the single-cell level. In glioblastoma, poor efficacy of receptor tyrosine kinase (RTK) therapies has been alternatively ascribed to genetic heterogeneity or to epigenetic transitions that circumvent signaling blockade.

**Results:** We combine cell lineage barcoding and single-cell transcriptomics to trace the emergence of drug resistance in stem-like glioblastoma cells treated with RTK inhibitors. Whereas a broad variety of barcoded lineages adopt a Notch-dependent persister phenotype that sustains them through early drug exposure, rare subclones acquire genetic changes that enable their rapid outgrowth over time. Single-cell analyses reveal that these genetic subclones gain copy number amplifications of the insulin receptor substrate-1 and substrate-2 (IRS1 or IRS2) loci, which activate insulin and AKT signaling programs. Persister-like cells and genomic amplifications of IRS2 and other loci are evident in primary glioblastomas and may underlie the inefficacy of targeted therapies in this disease.

**Conclusions:** A method for combined lineage tracing and scRNA-seq reveals the interplay between complementary genetic and epigenetic mechanisms of resistance in a heterogeneous glioblastoma tumor model.

**Keywords:** Therapy resistance, Glioma stem cells, Epigenetic, Genetic, Tumor heterogeneity, Lineage tracing, Single-cell RNA-seq, Insulin receptor substrate/IRS



## Background

Tumors comprise heterogeneous mixtures of cells that vary in their genetic makeup and epigenetic states. Outgrowth of fit genetic subclones is a well-established mechanism of drug resistance [1] and recent work has highlighted that genetic subclones may rapidly adapt through dynamic alterations involving extrachromosomal DNA [2, 3]. Moreover, transient epigenetic changes or cell state transitions that allow tumor cells to persist through drug exposure have been described in several cancer models [4–7]. The relative contributions of these respective models and how they cooperate to confer drug resistance in cancer remains a critical question (Fig. 1a).

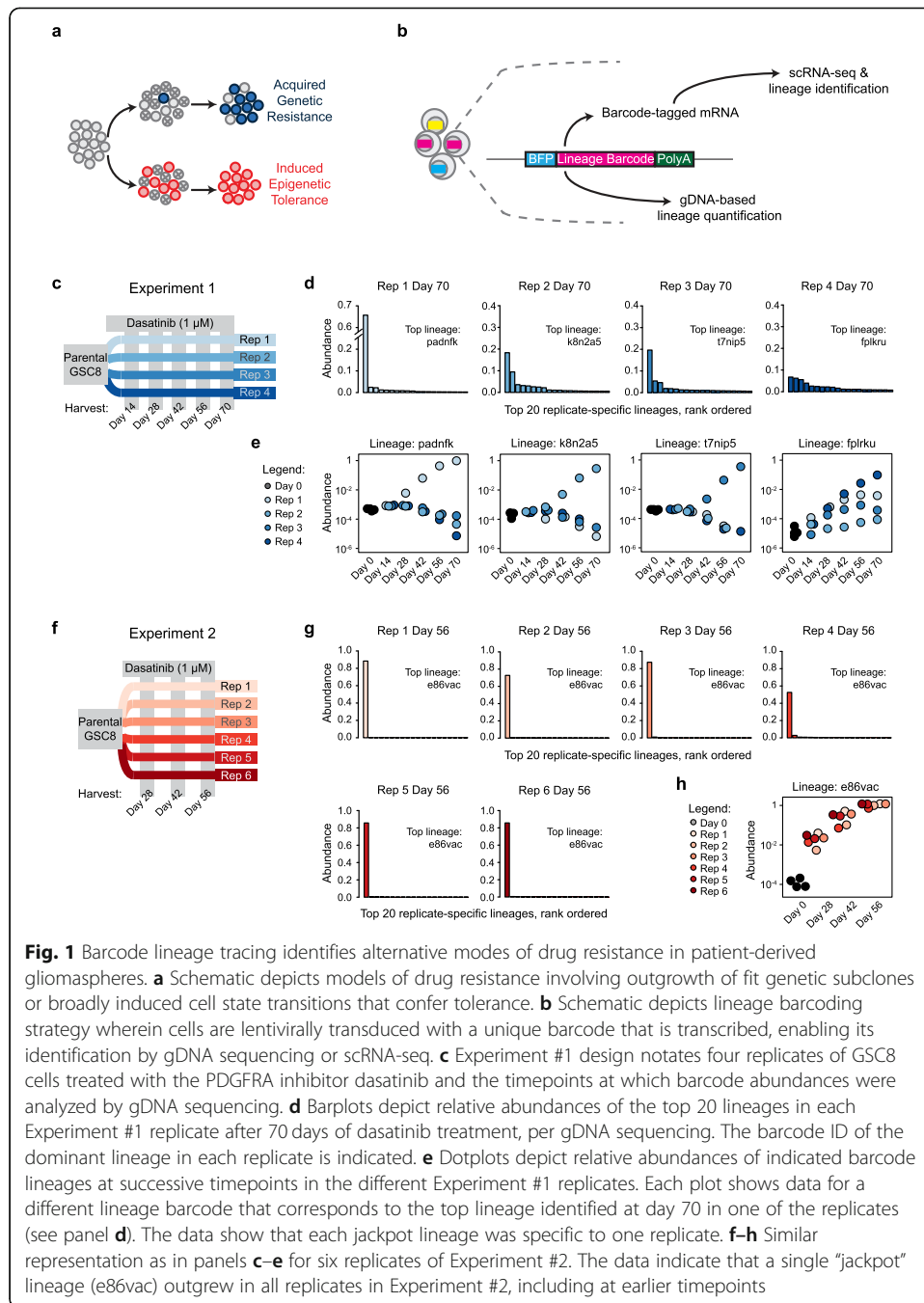
Genetic subclones and heterogeneous DNA methylation patterns within tumors have been characterized by sequencing-based methods [8–10]. Outgrowth of preexisting drug-resistant subpopulations has been tracked in experimental tumor models using DNA barcodes [11]. Epigenetic modes of treatment resistance have also been inferred [12, 13]. However, holistic assessment of preexisting and dynamic epigenetic and genetic drug resistance mechanisms remains an important goal that requires new methods.

Treatment resistance is of paramount concern in GBM, which inevitably relapses after treatment [14, 15]. Although these tumors frequently harbor RTK amplifications, therapies targeted against RTKs have failed in the clinic [16]. GBMs are fueled by stem-like subpopulations that tend to be relatively resistant to therapy [17] and can be modeled experimentally as gliomaspheres [18]. We therefore sought to develop strategies to evaluate the interplay between genetic and epigenetic mechanisms of RTK inhibitor resistance in gliomaspheres. We focused on a PDGFRA-amplified gliosphere model in which a subset of cells is able to withstand PDGFR inhibition by adopting a slow-cycling persistor state that is Notch-dependent and reversible [7]. We hypothesized that longer term dasatinib-resistant cell populations might depend upon both epigenetic persistor phenotypes and acquired genetic events. We developed, optimized, and employed a combined single-cell RNA-seq (scRNA-seq) and lineage tracing approach to investigate.

## Results

To track outgrowth and phenotypes of clonal lineages, we combined a lentiviral transgene barcoding system [11] with single-cell RNA sequencing (scRNA-seq). A diverse library of DNA barcodes was subcloned into the 3' UTR of a blue fluorescent protein transgene such that the barcode would be transcribed and captured in scRNA-seq data (Fig. 1b). For ease of identification, all barcode nucleotide sequences were converted to unique lineage IDs using Base32 encoding (Supplementary Fig. S1a). Deep sequencing confirmed that our starting library contained > 1 million barcodes (Supplementary Fig. S1b, see "Methods").

To maintain high barcode diversity in a patient-derived gliosphere model (GSC8), we transduced cells with lentiviral particles bearing this library at low multiplicity of infection (Supplementary Fig. S1c). Barcode diversity was determined after targeted sequencing of genomic DNA prepared from stably transduced cells, and reads were filtered to remove putative sequencing errors (Supplementary Fig. S1d). A homogenous baseline distribution was confirmed, with each barcode accounting for < 0.11% of the total reads (Supplementary Fig. S2a, Supplementary Table S1). The starting cellular barcode diversity after lentiviral infection and puromycin selection was on average 36,



**Fig. 1** Barcode lineage tracing identifies alternative modes of drug resistance in patient-derived gliomaspheres. **a** Schematic depicts models of drug resistance involving outgrowth of fit genetic subclones or broadly induced cell state transitions that confer tolerance. **b** Schematic depicts lineage barcoding strategy wherein cells are lentivirally transduced with a unique barcode that is transcribed, enabling its identification by gDNA sequencing or scRNA-seq. **c** Experiment #1 design notates four replicates of GSC8 cells treated with the PDGFRA inhibitor dasatinib and the timepoints at which barcode abundances were analyzed by gDNA sequencing. **d** Barplots depict relative abundances of the top 20 lineages in each Experiment #1 replicate after 70 days of dasatinib treatment, per gDNA sequencing. The barcode ID of the dominant lineage in each replicate is indicated. **e** Dotplots depict relative abundances of indicated barcode lineages at successive timepoints in the different Experiment #1 replicates. Each plot shows data for a different lineage barcode that corresponds to the top lineage identified at day 70 in one of the replicates (see panel **d**). The data show that each jackpot lineage was specific to one replicate. **f-h** Similar representation as in panels **c-e** for six replicates of Experiment #2. The data indicate that a single “jackpot” lineage (e86vac) outgrew in all replicates in Experiment #2, including at earlier timepoints

580 and 84,926 lineage barcodes across replicates of each experiment, respectively (Supplementary Table S1). Thus, in comparison to prior combined lineage tracing/scRNA-seq strategies used to study organismal development [19–22], our method harnesses a highly diverse set of barcodes and is uniquely powered to detect very rare subclones in malignant cell populations [23].

GSC8 gliomaspheres harbor a PDGFRA amplification and are largely sensitive to the PDGFRA inhibitor dasatinib but a subset are able to adopt a drug persister phenotype to survive initial drug treatment [7]. To ascertain mechanisms of drug persistence and resistance, in a first experiment we distributed barcoded GSC8 cells across four

separate cultures that were treated with dasatinib continuously over a 70-day time course. Aliquots of viable cells were collected from each replicate at days 14, 28, 42, 56, and 70, and barcode distributions were assessed by targeted DNA sequencing (Experiment #1; Fig. 1c). After 70 days of dasatinib exposure, dominant clonal lineages outgrew in three of the four replicates. These dominant “jackpot” lineages made up between 18.2 and 65.5% of these replicates, representing a >2000-fold relative expansion over the course of the experiment (Fig. 1d, Supplementary Fig. S2b; Supplementary Table S1). Their expansion was also evident in barcode distribution data for earlier time points (d42, d56) (Fig. 1e). Analysis of the fourth replicate did not reveal a jackpot lineage, but rather a spectrum of persistent lineages, all with less than 6.7% abundance at day 70 (Fig. 1d, Supplementary Fig. S2c). Notably, each of the three jackpot replicates was dominated by a different lineage. Moreover, the jackpot lineages were replicate-specific and did not show preferential fitness in any of the other replicates (<0.012%; Fig. 1e). Control DMSO-treated populations grown from the same parental populations did not exhibit jackpot clones (Supplementary Fig. S2d). Furthermore, none of the jackpot lineages from dasatinib-treated replicates 1–3 were within the top lineages detected in DMSO-treated cells (Supplementary Fig. S2e). This implied that the exceptional fitness of the dominant lineages in dasatinib-treated cultures was conferred by distinct, replicate-specific events that occurred after splitting of the barcoded population.

We repeated this experiment by transducing a separate parental culture of GSC8 gliomaspheres with cellular lineage barcodes (Experiment #2; Fig. 1f). After confirming a homogenous baseline distribution of diverse barcodes (Supplementary Fig. S2a), we aliquoted the barcoded cells across six replicates that were exposed to dasatinib as in the prior experiment. After 56 days of drug exposure, all six replicates contained a jackpot lineage that comprised between 52.4 and 88.5% of the population (Fig. 1g, Supplementary Fig. S2f). In contrast to the prior experiment, we found that all replicates were dominated by the same lineage (lineage barcode ID e86vac). This lineage was evident as a high abundance lineage as early as the day 28 timepoint (0.4 to 2.8%, Fig. 1h). However, it was not over-represented at the starting timepoint prior to dasatinib treatment (<0.014%). These results suggested that a single event incurred prior to splitting of the barcoded population conferred the exceptional fitness of the dominant lineage shared across these replicates.

We interpreted these lineage barcoding experiments as being representative of alternate resistance models: (1) an acquired resistance model could explain the distinct jackpot lineages evident at later timepoints in Experiment #1, (2) a preexisting resistance model could explain the single jackpot lineage shared across replicates in Experiment #2, and (3) an induced model could explain the persistence of a spectrum of lineages through the full course of Experiment #1 Replicate #4, as well as at earlier timepoints in all Experiment #1 replicates. We hypothesized that the induced resistance could reflect an epigenetic persistence mechanism, such as that we and others have described previously [6, 7, 24–26], while the instances of acquired and preexisting resistance might reflect more stable genetic alterations.

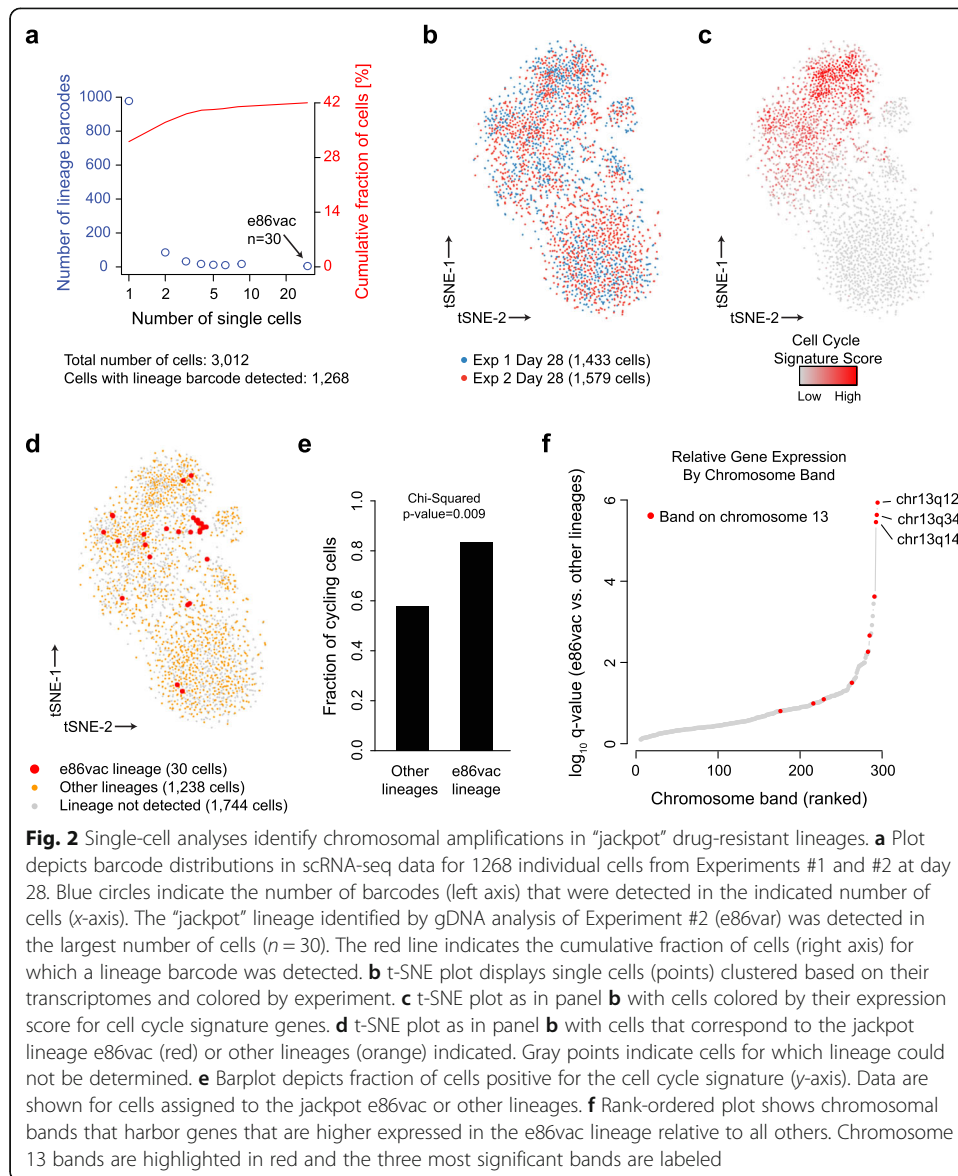
We therefore investigated the underlying alterations and mechanisms. First, we performed scRNA-seq on barcode-bearing gliomaspheres that we had collected from

Experiments #1 and #2 at day 28, reasoning that these data might reveal discriminating early features of the shared resistant subclone amongst a persister-rich background. We used the nanowell-based Seq-well technology [27] to acquire 3012 high-quality single-cell transcriptomes across both experiments at this timepoint. We were able to assign a unique lineage barcode to over 42.1% of cells with high confidence based on single-cell sequencing reads for the transcribed barcode. We identified a total of 1076 different lineage barcodes across the 1268 single cells with assigned barcodes (Fig. 2a). The most abundant lineage in the scRNA-seq data ( $n = 30$ , 1.9% of cells from Experiment #2) harbored the same barcode ID (e86vac) as the jackpot lineages identified in the DNA sequencing analysis of the six Experiment #2 replicates (Supplementary Fig. S3a). However, the highly-represented lineages from later timepoints in Experiment #1 were not over-represented at day 28, consistent with observations from targeted DNA sequencing that jackpot lineages from this experiment were not yet dominant at this early timepoint.

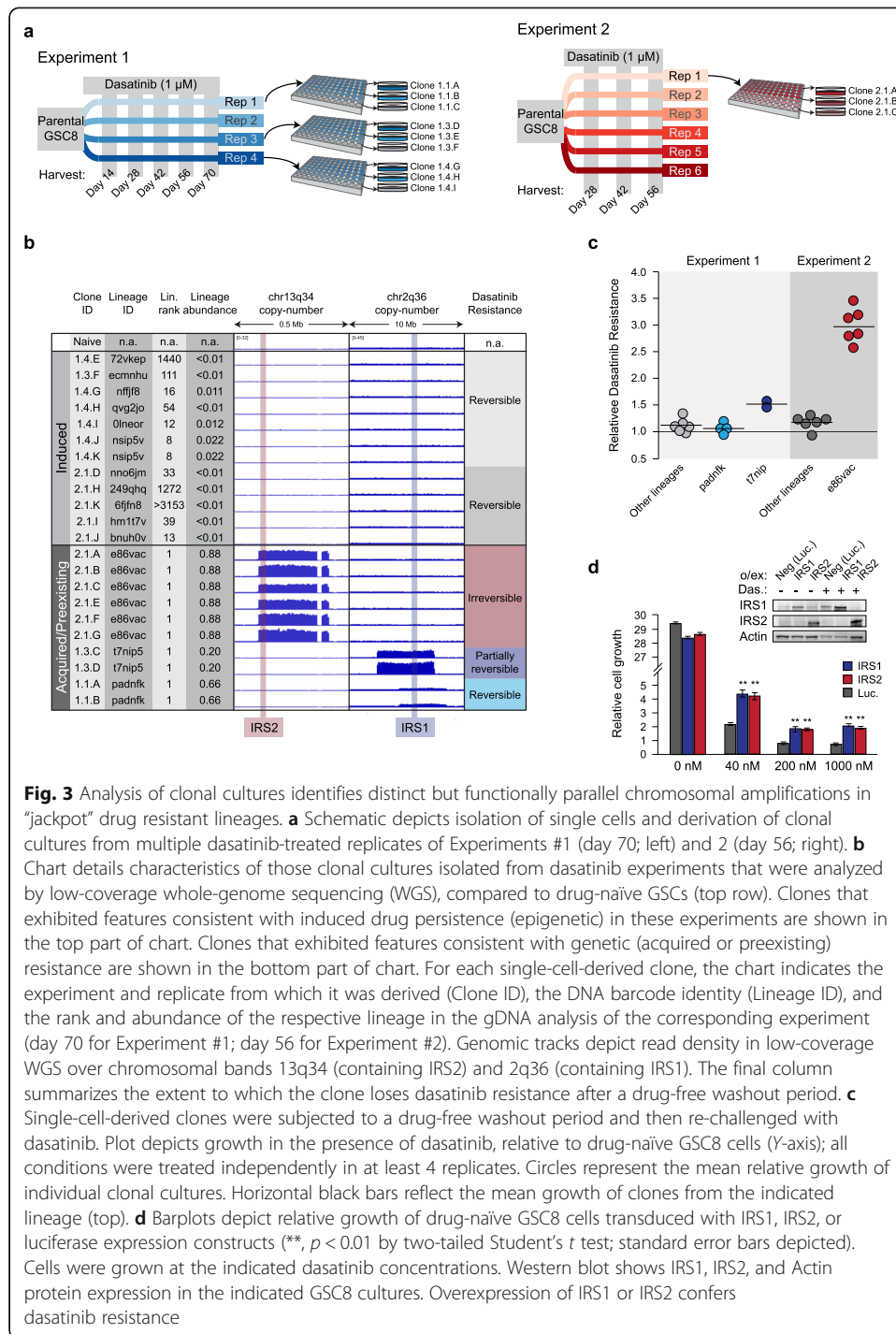
We first analyzed the transcriptional programs of individual cells by t-SNE visualization. This revealed strong overall concordance between the two experiments, both of which contained similar distributions of single-cell transcriptomes at day 28 (Fig. 2b). The t-SNE visualization revealed two main clusters of cells distinguishable by their expression of cell cycle genes (Fig. 2c, Supplementary Fig. S3b). Cells deriving from the predominant Experiment #2 lineage (e86vac) were enriched within the cell cycle-high cluster (Fig. 2d, e, Supplementary Fig. S3c), suggesting that they already had a proliferative advantage at early timepoints.

Since the e86vac lineage was the dominant lineage across multiple replicates, we hypothesized that it harbored a discrete genetic event such as a copy number alteration. To investigate mechanisms underlying the dasatinib-resistant phenotype of the dominant e86vac lineage, we aggregated single-cell transcriptomes for all cells matching this lineage and compared the expression profiles of e86vac cells against data aggregated from a cell-cycle matched cohort of cells from other lineages. We noticed that a number of upregulated genes in the e86vac lineage (e.g., COL4A1/2, SOX1, IRS2) reside within a single chromosomal band, chr13q34 (Supplementary Fig. S3d). Systematic evaluation of all chromosomal bands across the genome revealed that three bands on chromosome 13 (chr13q12, chr13q14, chr13q34) score most highly in e86vac lineage cells (Fig. 2f). We hypothesized that these transcriptional patterns reflect genetic copy number alterations, which are frequent in glioblastoma and other tumors [28, 29]. In particular, a potential role for focal chr13q34 amplification in RTK inhibitor resistance was of particular interest as recurrent amplifications of this locus have been reported in hepatobiliary, colorectal, breast, and rhabdomyosarcoma cancers [30–34]. We therefore sought to confirm a genetic alteration at this locus and its relationship to RTK inhibitor resistance.

To this end, we isolated clones from a variety of lineages identified in the dasatinib time course experiments. We flow sorted single cells from the final timepoints of both experiments into wells and expanded > 300 single-cell-derived cultures (Fig. 3a). Targeted sequencing of 75 of these cultures enabled us to assign unambiguous lineage barcodes (38 unique barcodes). These included the e86var lineage from Experiment #2 (24 clones), two jackpot lineages from Experiment #1 (padnfk and t7nlp5; 6 and 2 clones, respectively), and 35 non-jackpot lineages from both experiments (43 clones). We next



carried out low-coverage whole-genome sequencing (WGS) of 10 single-cell-derived jackpot lineages as well as 12 non-jackpot barcodes (Fig. 3b, Supplementary Table S2). Systematic analysis of copy number alterations across all 22 clones relative to parental GSC8 gliomaspheres revealed two prominent amplifications specific to the three jackpot lineages (Supplementary Fig. S4a-b). In line with our hypothesis, all e86var clones harbored a high-level amplification of chr13q34 (Fig. 3b; Supplementary Fig. S4c). Interestingly, the two jackpot clones isolated from Experiment #1 (padnfk and t7nlp5) did not exhibit any copy number alterations at chr13q34, but instead contained distinct, though overlapping, amplifications at chr2q36 (Fig. 3b; Supplementary Fig. S4b,c). In contrast, we did not detect any focal high-level amplifications in the non-jackpot clones, which were nonetheless persistent and viable after 56–70 days of dasatinib exposure (Fig. 3b). All clones maintained critical copy number alterations seen in the parental cells (e.g., PDGFRA, NMYC, MDM2; Supplementary Fig. S4a,c).



These results suggest that, in addition to inducing a known epigenetic persister intermediate population [7], dasatinib treatment of PDGFRA-amplified GSCs can prompt outgrowth of subclonal populations with focal amplifications of chr13q34 or chr2q36. Together, these varied mechanisms of treatment response suggest that cell populations from the same patient-derived gliomaspheres may adapt to targeted RTK therapy via multiple genetic and epigenetic mechanisms. We reasoned that the chr13q34 amplification evident in e86var likely represented a relatively stable event as it was present

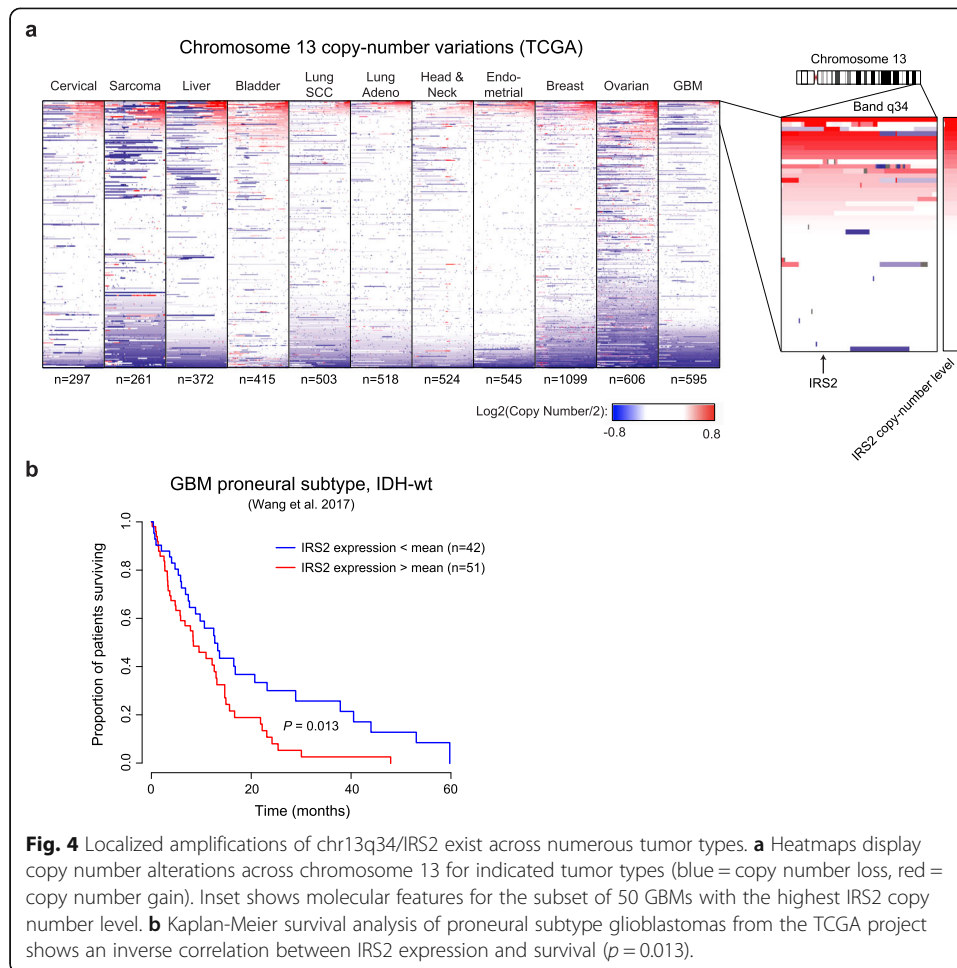
across all six replicates in Experiment #2. Indeed, we found that despite enduring dasatinib-induced inhibition of PDGFRA phosphorylation (Supplementary Fig. S4d), e86var clonal isolates cultured in the absence of dasatinib for > 4 weeks retained their drug-resistant phenotype when re-exposed to dasatinib (Fig. 3c). The chr2q36 amplified clones that arose differentially in Experiment #1 replicates were more variable and displayed some degree of drug resistance reversibility: clonal isolates with high copy number amplifications retained more stable dasatinib resistance than isolates with low copy number (Fig. 3c). In contrast, non-jackpot clones lost their drug tolerant phenotype entirely when cultured in the absence of dasatinib, consistent with a reversible epigenetic resistance mechanism.

To explore the mechanism by which GSC8 gliomaspheres acquire dasatinib resistance, we further investigated genes from chromosomal band chr13q34 that were upregulated in the e86vac jackpot lineage (Supplementary Fig. S3d). One of these genes was insulin receptor substrate 2 (IRS2; Fig. 3b), which has previously been identified as a low-frequency amplified gene in GBM [35] and is described as a putative driver oncogene in several other cancers [30, 32, 36–39]. Consistently, drug-naïve GSC8 gliomaspheres in which IRS2 was overexpressed exhibited robust dasatinib resistance (Fig. 3d). When we examined copy number data from the Cancer Genome Atlas [29, 35], we found that the chr13q34 locus including IRS2 was amplified in a subset of primary glioblastomas as well as multiple other primary tumor types (Fig. 4a). Kaplan-Meier survival analysis of samples from the IDH1-wildtype, proneural subtype of GBM [29, 40, 41], which most closely reflects the GSC8 model used here [41, 42], indicated that IRS2 overexpression was associated with poor patient prognosis (Fig. 4b). We observed this correlation within the entire subtype-filtered cohort (very few patients are annotated as having received RTK inhibitors), suggesting that IRS2 may serve an oncogenic role even in the absence of targeted treatments.

Remarkably, the chr2q36 amplifications observed in the jackpot lineages padnfk and t7nip5 from Experiment #1 contained the IRS2 paralogue insulin receptor substrate 1 (IRS1; Fig. 3b, Supplementary Fig. S4b,c) [43]. IRS1 has also been implicated in cancer initiation and progression [39, 44]. We therefore hypothesized that these paralogous members of the IRS protein family both confer resistance to RTK inhibitors. Indeed, overexpression of IRS1 in our GSC model also conferred dasatinib resistance (Fig. 3d). Combined with the observation that chr13q34 (containing IRS2) and chr2q36 (containing IRS1) were the two clear dominant high-level amplifications seen in jackpot clones relative to parental cells (Supplementary Fig. S4b), these results strongly suggest that amplification of the paralogous IRS1/2 genes provide outgrowth advantage for GSC8 cells under dasatinib treatment.

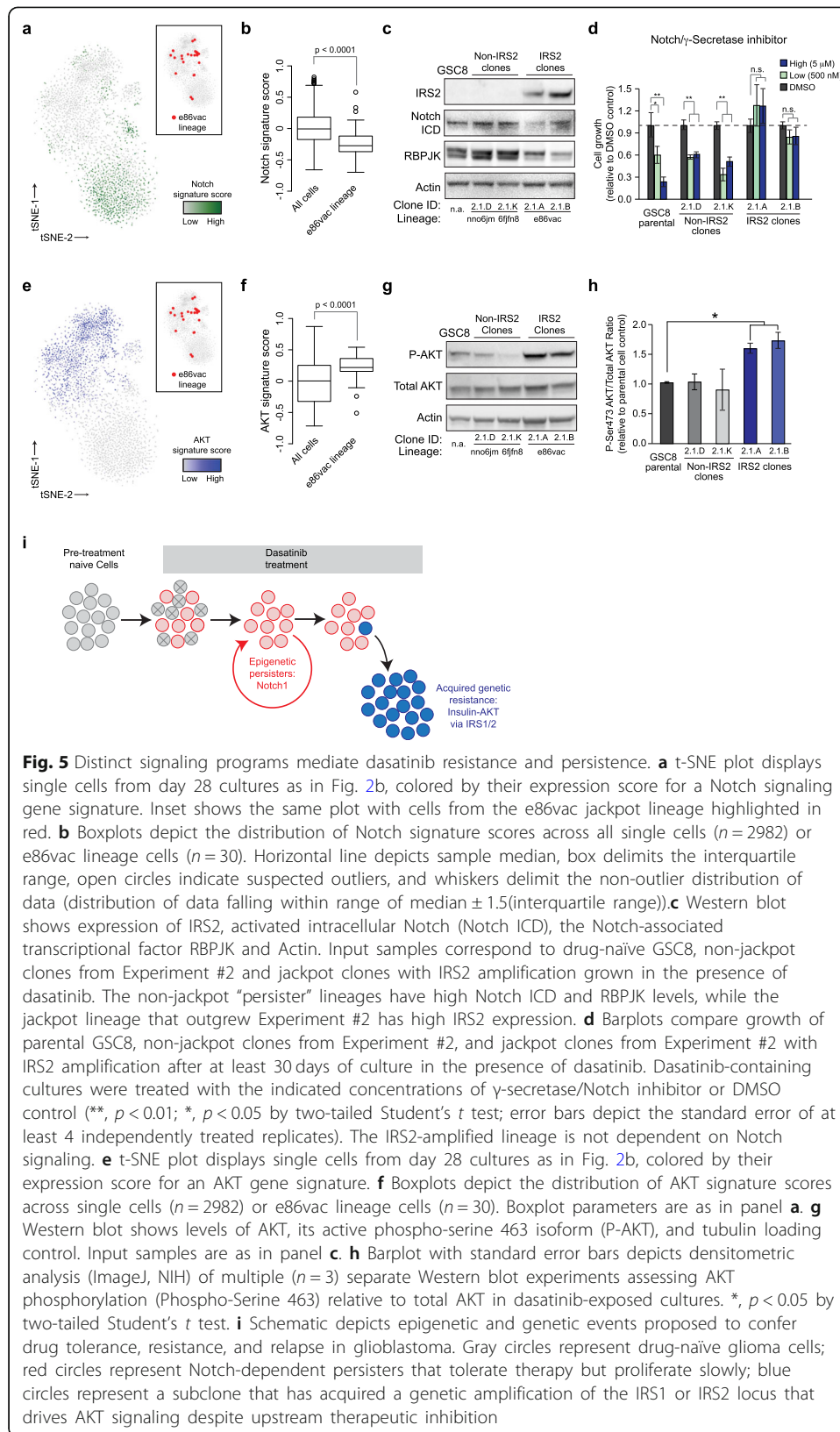
Finally, we sought to characterize the mechanisms that underlie dasatinib resistance in glioblastoma. We previously documented a critical role for Notch signaling in sustaining a dasatinib-tolerant slow-cycling persister state [7]. Consistently, scRNA-seq data for the dasatinib-treated gliomaspheres revealed subsets of cells that expressed a Notch gene signature (Fig. 5a) and a persister-like gene expression signature (Supplementary Fig. 5a). Notably, these clusters were depleted of cells from the IRS2-amplified lineage e86vac, which had low Notch signature gene expression and low persister signature gene expression (Fig. 5a, b, Supplementary Fig. S5a). Further, whereas dasatinib-treated GSC8 cultures and non-jackpot lineages expressed high levels of Notch





intracellular domain (ICD) and downstream transcription factor RBPJK, the IRS2-amplified lineage demonstrated decreased evidence of active Notch signaling (Fig. 5c). Consistently, persister-rich GSC8 populations were sensitive to gamma secretase inhibition of the Notch pathway, while the IRS2-amplified clones were largely insensitive to Notch pathway inhibition (Fig. 5d). Thus, IRS2-mediated dasatinib resistance is independent of Notch.

IRS1 and IRS2 are both highly regulated adaptor proteins that link insulin receptor (IR) or insulin-like growth factor (IGF-1R) signaling to downstream effector pathways such as AKT and ERK (schematic in Supplementary Fig. S5b) [45]. Further analysis of the scRNA-seq data from Experiment #2 revealed an increased AKT expression signature within the IRS2-amplified lineage, relative to all other lineages (Fig. 5e,f). Western blots confirmed that in dasatinib-containing conditions, IRS2-amplified clonal lines demonstrated increased AKT activation (p-AKT) relative to parental cells and control clonal lines (Fig. 5g,h). In contrast, ERK appeared largely inactive in dasatinib-treated cells and its activation was uncorrelated with IRS2 status (Supplementary Fig. S5c). Further, IRS2-amplified clones lost sensitivity to MEK/ERK pathway inhibition (Supplementary Fig. S5d). Nonetheless, the amplified clones were highly sensitive to inhibitors of IGF-1R/IR signaling (Supplementary Fig. S5e), supporting a prominent role for IGF-1R/IR/IRS/AKT signaling in their ability to rapidly proliferate in the absence of PDGFRA signaling.



## Discussion

Here, we combined a high complexity lineage barcoding system with scRNA-seq to investigate the mechanisms by which PDGFRA-dependent stem-like GBM cells evolve resistance to PDGFRA inhibition. The data reveal alternate resistance mechanisms involving genetic or epigenetic alterations (Fig. 5i). A stable IRS2 locus amplification that enables rapid growth in the presence of dasatinib may preexist in patient-derived gliomaspheres and has been documented in primary GBMs and other tumors prior to treatment [35] (Fig. 4a). More variable copy number alterations of the IRS1 locus also facilitate rapid growth and may arise de novo during drug exposure. Finally, a rapidly reversible persister state sustains viability of a subset of cells by epigenetic mechanisms. Notch-positive persister-like cells are already present in primary tumors [7] and may bridge GBM cells through early treatment until fully drug-resistant genetic alterations arise.

These results are an important foundation for testing the hypothesis that epigenetic persistence and genetic amplifications may jointly contribute to the development of resistance to RTK inhibitors or other targeted therapies. Epigenetic persisters have been characterized as a flexible, reversible phase of therapy response [4–7]. Here, we further note that genetically distinct subclones bearing IRS1 amplifications also exhibit a degree of reversibility to dasatinib resistance that may depend on copy number and yet undetermined factors. Thus, certain genetic alterations may provide another layer of dynamic adaptability in cancers, as noted in recent work profiling rapid adaptation via extrachromosomal DNA [2, 3, 46]. Further characterization of the multiple layers of epigenetic and genetic flexibility in models of treatment response could aid in the rational design of therapeutic approaches that address both the reversible and irreversible phases of therapeutic response.

Evaluation of clinical samples from completed and ongoing clinical trials [47–49] should be prioritized to determine the contribution of IRS1/2 expression to therapy responsiveness. Additionally, prospective evaluation of the contribution of IRS1/2 overexpression to patient outcomes and tumor aggressiveness in the absence of RTK treatment could further elucidate the mechanistic specificity of these amplifications in patient tumors.

## Conclusions

Combining high complexity barcoding with single-cell RNA-seq permitted the identification of IRS2 and IRS1 copy number amplifications as novel genetic mediators of dasatinib resistance in a PDGFRA-amplified gliomasphere model, in addition to transcriptionally characterizing an intermediate Notch-high epigenetic persister population. Our study provides important insights into how these alternate modes of drug tolerance and drug resistance can conspire to overcome therapy and cause the inevitable relapse of this disease.

## Methods

### Cell models and culture

Surgical GBM specimen-derived neurosphere cultures (GBM8 a.k.a. GSC8) were isolated and previously characterized [50]. Based on gene expression and mutational analysis this model most closely resembles an IDH1-wildtype, PDGFRA-amplified

glioblastoma of the proneural subtype [41, 42]. Cells were maintained in stem cell-permissive, serum-free, non-adherent spheroid cultures using Neurobasal medium (Gibco) supplemented with N2/B27 (Gibco), penicillin/streptomycin (Gibco), Gluta-MAX (Gibco), recombinant human EGF (20 ng/mL, R & D systems), recombinant human FGF2 (20 ng/mL, R&D systems), and heparin (2000 ng/mL, Sigma). Cell dissociation was accomplished with trituration or Accutase (BD Bioscience) treatment.

### **Barcode plasmid library construction**

Plasmid pBA439 was a gift from Jonathan Weissman (Addgene plasmid # 85967) [51] and was modified to remove the mU6 promoter by NotI and HpaI restriction digest and self-ligation after Klenow (NEB M0210S) treatment. The resultant pBA439HN modified plasmid was used as the backbone for generating the barcode library.

To generate the semi-random barcoded insert, synthetic oligos oligoF (ATGCCG TCTCCCTAGGACTGACTGCAGTCTGAGTCTGACA GWSWSWSWSWSWSWSWSWSWSWSWSWSWSWSWSWSWSAGCTACGCACTCTATGC-TAGTGCTAG; W = adenine or thymine and S = guanine or cytosine) and oligoR (gcatcgctcGAATTCCTAGCACTAGCATAGAGTGGGTAGCT) were purchased from IDT. This WSx15 semi-random sequence is referred to as the “lineage barcode.” OligoF and oligoR were annealed and the double-stranded oligonucleotides were generated by extension reaction with Phusion High-Fidelity DNA polymerase (NEB). Compatible ends were generated by treating the pooled duplex barcoded insert and the vector with AvrII and EcoRI. The pooled library was then subcloned into restriction-digested vector pBA439HN. Restriction enzyme-treated pBA439HN and the barcode double-stranded oligos were mixed at a ratio 1:4 vector:insert with  $\sim 4 \times 10^{11}$  digested vector molecules, ligated with T4 DNA ligase (NEB) at 16 C and ligated products were purified by DNA clean & concentrator 5 kit (Zymo research). About 25% of the purified ligated products were transformed into ElectroMAX Stbl4 cells (Thermo Fisher Scientific), and cells were plated on ampicillin/LB agar plates. All plated colonies were pooled and collected, and the resultant plasmid library was purified with HiSpeed Plasmid Maxi Kit (Qiagen). The purified plasmid library was sequenced and reads filtered as described below showing a diversity of 1.056 million barcodes with 2 or greater reads/barcode.

### **Lentiviral-mediated stable lineage barcoding of GBM cells**

For stable incorporation of the barcoded lineage construct, pooled pBA439 library plasmid was cotransfected with psPAX2 (Gift of Didier Trono, Addgene plasmid # 12260) and VSV. G plasmids (Gift of Tannishtha Reya, Addgene plasmid # 14888) at a ratio of 3: 2:1 into 293T cells using FuGENE HD (Promega) per manufacturer’s instructions. Media was changed after 6 h, and cells were cultured for 3 days. Then viral supernatant was collected and concentrated using Lenti-X concentrator (Clontech) following the manufacturer’s instructions, and an estimated MOI for the viral preparation was determined with serial dilution treatment of GSCs and evaluation for tagBFP expression by flow cytometry. For each replicate, approximately  $1 \times 10^6$  GSCs plated in stem cell-permissive monolayer culture with laminin (5  $\mu$ g/mL Engelbreth-Holm-Swarm laminin, Sigma) were infected

with concentrated virus at an MOI of 10% for 24 h prior to selection of barcode-incorporated GSCs using puromycin (0.5  $\mu\text{g}/\text{mL}$ , Life Technologies) for 72 h.

#### **Dasatinib treatment and passaging**

Equivalent numbers ( $1 \times 10^6$ ) of stably transduced barcoded cells were plated in replicates and treated with 1  $\mu\text{M}$  dasatinib. Dissociation and passage of cells with renewal of compound-containing cell culture medium was performed at least weekly. Each plate was passed and treated independently without cross-contamination of replicates.

#### **Isolation, amplification, and sequencing of genome-derived barcodes**

Following previously reported protocols [11], dasatinib-treated cells bearing the lineage barcode library were harvested and counted at indicated experimental time points. Genomic DNA (gDNA) was prepared with a QIAmp DNA blood mini Kit (Qiagen). The lineage barcode was amplified for multiplexed next-generation sequencing (NGS) using primers bearing Illumina adaptor and index sequences. PCR amplification of barcode was performed on half of the gDNA isolated for each replicate and time point. PCR amplification was performed in one or more reactions to ensure the equal amplification of barcodes; up to 2  $\mu\text{g}$  of genomic DNA was used as a template for each PCR reaction with Titanium Taq DNA polymerase (Clontech) per manufacturer's instructions. Amplified DNA was pooled if multiple PCRs were performed for a given sample, and samples were purified with QIAquick PCR purification columns (Qiagen). TapeStation HS D5000 (Agilent) analysis was used to verify the purity and size of the indexed PCR product and quantification with Qubit dsDNA HS (Thermo Fisher) was performed prior to multiplexed NGS Illumina sequencing using a NextSeq500 instrument (Illumina). Read 1 was sequenced 38 bp (or longer) with a custom spike-in primer (CACTGACTGCAGTCTGAGTCTGACAG) with 8 bp i7 sequencing read. PhiX was added as about 10% of the total reads. Between 2.8 and 10.2 million reads were generated for each library (Supplementary Table S1). Amplification and sequencing of the barcode plasmid library was performed in a similar way.

#### **Computational analysis of genome-derived barcodes**

Fastq files from barcode sequencing libraries were assessed for high-quality barcode reads by filtering to include only those reads that met the expected WSx15 pattern (see "[Barcode Plasmid Library Construction](#)" for details on semi-random barcode design). This sequence pattern was identified in at least 86.4% of reads in all samples. Most reads containing sequencing errors (including single base insertions or deletions) are filtered during this step as they do not match the expected pattern. This initial filtering step was performed using the `grep` unix command. Further analysis on summarized barcode sequences was performed using the R programming language [52]. We employed a computationally efficient method to filter out additional sequences that likely contained sequencing errors. For each sequence, we identified all 30 possible sequences with a single base mismatch (i.e., a Hamming distance of 1). If one of these 30 sequences was observed more frequently than the original sequence, the original sequence was filtered out as it likely reflects a sequencing error. Notably, this approach does not only remove single base sequencing errors, but also multiple base sequencing errors, which are in turn less frequent than single base errors (illustrated in Supplementary Fig. 1d). After filtering,

between 2,766 and 88,469 unique barcodes were detected for different samples ( $\geq 10$  sequencing reads), greatly varying in abundance (top lineage abundance between 0.1 and 88.5%; detailed information on each sample is provided in Supplementary Table S1). Estimation of the underlying barcode complexity in the plasmid library was performed in a similar way, requiring at least two sequencing reads per barcode.

### Single-cell transcriptome profiling using Seq-well

Nanowell-based single-cell transcriptome isolation, indexing, and sequencing were performed as previously described [27, 53]. Briefly, Accutase-dissociated and filtered GSCs were counted and 10,000 single cells were loaded onto a Seq-Well nanowell array of ~90,000 wells with pre-loaded beads bearing oligonucleotides containing a primer-binding sequence, cell barcode identifier, UMI, and oligo-dT sequence (Chemgenex). The array was sealed with a partially permeable polycarbonate membrane (Sterlitech Custom Order), cells were lysed in the sealed wells, and mRNA were hybridized to the oligonucleotide-conjugated beads. After liberating and pooling all beads from the array, reverse transcription was performed with Maxima H-RT (Thermo) and a template switching oligo (AAGCAGTGGTATCAACGCAGAGTGAATrGrG+G, Exiqon). Exonuclease treatment with exonuclease I (NEB) was performed prior to whole transcriptome amplification (WTA) using 2x KAPA HiFi Hotstart Readymix (KAPA) and SMART PCR primer (AAGCAGTGGTATCAACGCAGAGT, IDT). AMPure SPRI bead cleanup (0.6X, Beckman Coulter) was performed and a Nextera XT tagmentation kit (Illumina) was used to prepare indexed sequencing libraries, with a P5-SMART-Hybrid primer (AATGATACGGCGACCACC GAGATCTACACGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGT\*A\*C) and a barcoded N70X primer for sample identification (CAAGCAGAAGACGGCATAACGAGATXXXXXXXXGTCTCGTGGGCTCGGAGATGT) from IDT. Libraries were then sequenced using a custom Read1 primer (GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC) on a NextSeq500 instrument (Illumina). The following specifications were used for paired-end sequencing: 20 bp for read 1 (containing 12 bp cell barcode information and 8 bp UMI), 8 bp for i7 index, and 64 bp for read 2 (containing part of the transcript). Fastq files for each library were generated using bcl2fastq allowing for up to one mismatch to expected library barcodes. Identification of cell barcodes, genome alignment, and quantification of gene expression levels for each single cell was performed as previously described [53] using the splice-aware aligner STAR version 2.6.0c [54] and the R programming language version 3.3.0 [52]. For downstream analyses, we retained high-quality cells with at least 2500 UMIs and normalized expression values to a total of 10,000 per cell before log-transformation (after addition of 1).

### Barcode enrichment from single-cell transcriptomes

A portion (20%) of the WTA material was utilized for PCR enrichment of lineage barcode-containing transcripts to efficiently link cell barcodes with lineage barcodes, as previously described [53]. PCR was performed using Titanium Taq Polymerase (Clontech) and the P5-SMART PCR Hybrid PCR and Biotin-R18 Oligo (/5Biosg/GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATAGAGTGGCTAGCT) from IDT. Cycling parameters were as follows: (1) 95 °C for 5 min; (2) 13–15 cycles of (a) 95 °C for 30 s, (b) 55 °C for 30 s, and (c) 72 °C for 30 s, (3) final extension with

72 °C for 2 min. Purified PCR products (Qiagen PCR Purification Kit) were incubated with Streptavidin C1 Dynabeads (Thermo Fisher) in 1×BW buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl) for 15 min at room temperature, washed with 1×BW buffer 5 times, and washed with EB buffer twice. Beads were resuspended in EB and used for PCR with P5-SMART-Hybrid and barcoded N70X primers, and Titanium Taq Polymerase with the same parameters as above. The liberated PCR products were separated from the beads and purified with 0.6X AMPure SPRI purification prior to quantification with Qubit dsDNA HS (Thermo Fisher) and quality assessment with HS D5000 TapeStation (Agilent). Following paired-end sequencing, reads were processed using the cell barcodes identified in the corresponding Seq-Well library, as described previously [53].

### Computational analysis of transcriptome-derived barcodes

Similar to genome-derived barcode libraries, sequencing reads from transcriptome-derived barcode libraries were assessed for high-quality lineage barcodes. Reads were first filtered to include only those reads that started with the target sequence of the biotinylated primer (CATAGAGTGCGTAGCT), followed by the WSx15 pattern. This initial filtering step was performed using the `grep` unix command. Further analysis on summarized lineage barcodes (annotated by cell barcode and UMI) was performed using the R programming language version 3.3.0 [52]. We filtered sequences according to the following five criteria: (1) The combination of cell barcode, UMI, and lineage barcode was detected by at least 20 reads. (2) The lineage barcode was identified in the genome-derived libraries of any of the four replicates at day 0 in the respective experiment. (3) The cell barcode matched a high-quality cell in the corresponding Seq-Well library. (4) For a given cell barcode/UMI combination, at least 95% of reads represented a single lineage barcode. (5) Similarly, for a given cell barcode, at least 95% of reads represented a single lineage barcode. Using these criteria, 1268 of 3012 high-quality single cells (42.1%) were matched to a defined lineage barcode.

### Clustering, signature scores, and gene set enrichment analysis

All subsequent computational analyses were performed using the R programming language version 3.3.0 [52]. For t-SNE visualization of single-cell expression profiles, we first determined variably expressed genes as previously described across both experiments ( $n = 427$ ) and centered the log-transformed expression values for each gene. We then used 1 minus the Pearson correlation coefficient as a distance measure between cells and used the `Rtsne` package version 0.16 [55, 56] with the following non-default parameters: `pca = F`, `is_distance = T`, `theta = 0`.

For deriving single-cell gene expression signature scores, we used previously reported gene signatures that exemplified cycling cells [57], Notch1 signature (Naive GSC8 N1ICD target genes) [7], persister signature (Cluster 4) [7], and AKT signaling [58]. Signature scores were calculated as described previously [53], scoring each gene relative to the 100 genes with the smallest difference in average expression level, and averaging over all genes in the signature. Since all signatures except the cycling signatures consisted of several hundreds of genes, we refined signatures by first calculating scores using all genes, and

then only retained the 50 most informative genes (identified by correlating gene expression levels with signature scores) and recalculated scores for the shortened signature.

Gene set enrichment analysis (GSEA) for chromosomal bands was performed in bulk mode using the Liger package version 1.0 [59, 60] using default parameters. For chromosomal bands with a minimal  $p$  value, we ran the analysis for a second time using  $10^6$  random permutations.

#### **Clonal culture isolation and characterization**

Populations of lineage-barcoded, dasatinib-treated cells from d70 (Experiment #1) and d56 (Experiment #2) were sorted using single-cell sorting parameters on a Sony SH800 flow cytometer. Viable cells were determined with Live/Dead near-IR staining (Invitrogen) or propidium iodide exclusion and single viable cells were sorted into U-bottom Ultralow Adhesion 96-well plates (Corning). Low clonal culture outgrowth was noted on initial attempts utilizing 100% Neurobasal with growth factor supplements, so subsequent isolations utilized 50% Neurobasal with growth factor supplements and 50% filtered conditioned media from GSC8 cells cultured in Neurobasal with growth factor supplements for 36 h with substantially higher success of clonal culture outgrowth. Clonal cultures were permitted to grow in dasatinib-free conditions with periodic dissociation and replating in serially increased volumes of fresh Neurobasal with growth factor supplements until sufficient cell numbers were available to viably freeze 2 aliquots of cells and harvest gDNA (DNeasy 96 Blood and Tissue Kit, Qiagen). gDNA was quantified with Nanodrop, and lineage barcode isolation and analysis was performed on samples with sufficient high-quality gDNA for analysis ( $> 10$  ng/ $\mu$ L). Barcode amplification was performed using Q5 2X High-Fidelity Mastermix (New England Biolabs) and the following primers obtained from IDT: ClonePCR\_Fwd: GATGCCTGGCGTCTACTATGT and ClonePCR\_Rev: CTGATCAGCGGGTTTAAACGG with the following cycling conditions: 1 cycle of 98 °C for 30 s; 35 cycles of 98 °C for 10 s then 67 °C for 30 s then 72 °C for 30 s; elongation for 2 min at 72 °C. Gel-purified PCR products (Qiagen Gel Purification Kit) were Sanger sequenced (Quintara Biology) using the ClonePCR\_Rev primer above. Clones were selected for further analysis based on sequencing quality and lineage representation, and selected clonal cultures were thawed and reassessed a second time by gDNA isolation and Sanger sequencing to confirm lineage.

#### **Low-coverage whole-genome sequencing**

Cells were snap-frozen, gDNA was isolated using DNeasy Blood and Tissue Kit (Qiagen), and then quality and quantity of gDNA was determined by Nanodrop. In total, 50 ng gDNA was used for fragmentation and sample preparation/amplification using the Nextera DNA kit (Illumina) and Nextera i7/i5 primers (Illumina) according to the manufacturer's instructions. AMPure-cleaned DNA was quantified with Qubit HS dsDNA kit (Thermo Fisher) and PCR product size and purity assessed with TapeStation D5000 (Agilent). Samples were diluted and mixed in equimolar quantity and paired-end sequenced ( $2 \times 38$  bp) in multiplexed fashion using a NextSeq500 instrument (Illumina), acquiring between 10.8 and 37.5 million reads for each library (detailed information on each sample is provided in Supplementary Table S2). Fastq files were generated using bcl2fastq and aligned to the human reference genome (hg19) using bwa mem (version 0.7.17). Alignments were



coordinate sorted and putative PCR duplicates were removed using samtools. Additionally, read pairs with an insert size larger than 1 kb were also removed. Read coverage in genomic bins of 50 kb was calculated using igvtools count (version 2.5.0). Resulting coverage tracks were subsequently normalized by scaling the median coverage of chromosome 5 to 3 (chromosome 5 did not show any apparent copy number variations between all samples). Samples were visualized using the IGV browser.

### Cell growth assays

The following inhibitors were used: dasatinib (Selleck, S1021), IR/IGFR1 inhibitor GSK1904529A (Selleck S1093), Notch/Gamma secretase inhibitor Compound E (Millipore 565790), and MEK/ERK1/2 pathway inhibitor PD98059 (Selleck S1177). Accutase-dissociated gliomaspheres were counted and plated at 1000–5000 cells/well of a 96-well plate in 125  $\mu$ L Neurobasal with growth factors with either DMSO or dasatinib and/or other targeted inhibitors in at least triplicate. A set of cells of each type were analyzed for total ATP content at day 0 for normalization. Cells were permitted to grow for 7 days before they were analyzed per manufacturer's protocol by Cell Titer Glo (Promega) alongside ATP standards for end point luminescence on a Synergy HTX Plater-reader (BioTek). At least 4 separately treated replicates were quantified for each condition.

### Resistance quantification by AUC analysis

Relative resistance to dasatinib was evaluated by comparing the drug-response survival curves of cell populations normalized to drug-naïve GSC8 cells at various concentrations of dasatinib ranging from 1.6 nM to 5  $\mu$ M. Comparative AUC analysis was performed using computeAUC in the PharmacGx package in R [52] and normalizing AUC values to that of drug-naïve GSC8 cells from the same experiment. A total of 12 clonal cultures from Experiment #1 and 12 from Experiment #2 were analyzed (including but not limited to the clones evaluated by WGS in Fig. 3b).

### Lentiviral-mediated overexpression of IRS1/2 in GBM cells

For stable incorporation of overexpression constructs, 10.5  $\mu$ g pSMAL plasmid [61] with control (pSMAL-luciferase) or IRS1/2 overexpression modules were cotransfected with 1.9  $\mu$ g REV, 3.8  $\mu$ g RRE, 2.7  $\mu$ g VSV-G, and 4.5  $\mu$ g pADV plasmids per 10 cm dish with FuGENE HD (Promega) per manufacturer's instructions. Media was changed 8 h after transfection, then harvested and concentrated at 48 h post-transfection. Approximate viral stock MOI was estimated by GFP positivity by flow cytometric analysis after serial dilution treatment of GSCs. GSCs were treated with equivalent MOIs of control (pSMAL-Luciferase) and pSMAL-IRS1 and pSMAL-IRS2 virus, and stable populations with similar GFP positivity were obtained by flow sorting on a Sony SH800S Flow Sorter and IRS overexpression levels were verified by Western blotting (Fig. 3d).

### Western blot analysis

Cells were treated with DMSO or dasatinib continually for at least 30 days and then harvested for analysis. PBS-washed, live cells were snap-frozen and lysed using RIPA buffer (Thermo Fisher) supplemented with HALT protease and phosphatase inhibitors

(Thermo Fisher) on ice for 20 min, lysates were homogenized by drawing through a 23-Gauge needle, and debris was removed by centrifugation. Protein content was quantified in triplicate by BCA assay (Thermo Fisher). Equivalent lysate quantities (25–30 µg/well) were brought to the same volume, prepared with 4X LDS sample buffer (Thermo Fisher) and TCEP reducing agent (Thermo Fisher) per manufacturer's instructions, and treated at 95 °C for 5 min. Samples and molecular weight ladder (BioRad Kaleidoscope Protein Ladder) were loaded on a NuPAGE 4–12% gradient bis-tris precast polyacrylamide gel (Thermo Fisher) and running at 90–125 V until the dye front passed from the gel. Protein was transferred at 20 V for 6 min using the iBlot transfer system (Thermo Fisher) onto prepared nitrocellulose iBlot transfer stacks. After blocking for 1 h in 5% BSA in TBST at room temperature, membranes were incubated in indicated concentrations of antibodies overnight at 4 °C in TBST+ 5% BSA. Antibodies were used at following concentrations: p-Ser473 AKT (Cell Signaling Technologies, CST #4060, 1:1000, lot 23), total AKT (CST #2920, 1:1000, lot 8), p-ERK1/2 (CST #4370, 1:1000, lot 24), total ERK (CST #9107, 1:1000, lot 10), cleaved Notch ICD (CST #4147, 1:1000, lot 6 and 7), RPBJK (Abcam ab25949, 1:2000, lot GR3240322-1), IRS2 (Abcam ab134101, 1:2000, lot GR219213-16), IRS1 (Abcam ab40777, 1:2000, lot GR278510-7). Membranes were extensively washed with TBST, incubated with secondary antibodies at 1:2500 (Goat anti Rabbit IgG Starbright Blue 700, Goat anti Mouse IgG Dylight 800, Biorad or HRP anti-Rb/Mo, CST #7074 and 7076) and 1:2500 rhodamine-conjugated FAB anti-actin or anti-tubulin (Biorad) for 1 h at room temperature, and then washed extensively with TBST. Fluorescently labels were visualized with GelDoc imager (BioRad). Membranes incubated with HRP-conjugated antibodies were treated for 1–5 min with ECL prepared according to the manufacturer's instructions (Millipore) prior to chemiluminescent imaging with GelDoc Imager (BioRad). Uncropped Western blot data, including ladders, are displayed in Supplementary Fig. S6.

#### **Analysis of tumor copy number and expression data from TCGA**

To generate tumor type-specific heatmaps of copy number data across tumor types, the Cancer Genome Atlas (TCGA) was accessed via UCSC's [xenabrowser.net](https://xenabrowser.net). TCGA data for each tumor type was sorted by IRS2 copy number level and chromosome 13 copy number data visualized using custom display settings (origin = 0, threshold = 0.3, saturation = 0.8).

Survival analysis of TCGA GBM samples was performed using processed exon array gene expression data obtained from the NCI GDC data portal (<https://portal.gdc.cancer.gov>). Sample annotation was obtained from the supplementary tables of the primary publication [29, 35]. We only included IDH-wildtype cases associated with the proneural subtype as identified by previously reported subtype-specific gene lists in IDH1 WT tumors [41]. Kaplan-Meier analysis was performed using the survival package in R, splitting cases in two equally sized groups based on their expression for IRS2.

#### **Supplementary information**

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02085-1>.

**Additional file 1.** Supplemental Figures and Legends (Figs. S1-S6). Contains compiled supplementary figures and legends referenced in the main text.

**Additional file 2.** Supplemental Table S1. Contains summary data (e.g., number of reads, number of barcodes, top barcode ID and prevalence, etc.) from each lineage barcode gDNA library sequenced and reported in the main text.

**Additional file 3.** Supplemental Table S2. Contains summary data for each low-pass WGS library sequenced and reported in the main text.

**Additional file 4.** Review history.

#### Acknowledgements

We thank J.A. Verga and R. Boursiquot for technical assistance and B. Liao, C. Sievers, F. Najm, L. Gaskell, and other Bernstein lab members for helpful discussions.

#### Review history

The review history is available as Additional file 4.

#### Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

#### Authors' contributions

C.E.E., H.M., and P.V.G. conceptualized and designed the experiments. C.E.E., H.M., and S.J.V. optimized and executed the experiments. V.H. performed the computational analysis. B.E.B. provided senior guidance. C.E.E. and B.E.B. wrote the manuscript with assistance from other authors, all of whom approved the final submission.

#### Authors' information

Present addresses (if different from title page):

C.E.E.: Department of Radiation Oncology, Duke University Hospital, Durham, NC, USA

H.M.: Oncology Research Laboratories II, Daiichi Sankyo Co., Ltd. Tokyo, Japan.

P.v.G.: Division of Hematology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

#### Funding

C.E.E. is supported by an appointed KL2 award from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award KL2 TR002542). The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

B.E.B. is the Bernard and Mildred Kayden Endowed MGH Research Institute Chair and an American Cancer Society Research Professor. This research was supported by the National Cancer Institute (DP1CA216873) and SCC award I9-A9-071.

None of the funding bodies were directly involved in the design of the study, nor in the collection, analysis or interpretation of the data, or writing of the manuscript.

#### Availability of data and materials

Sequencing reads and processed data that supports the findings of this study have been deposited in GEO with the accession code GSE142119 [62]. To access data, go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142119>.

The authors declare that data processing steps are described in detail within the methods such that analysis can be replicated.

The authors declare that all other data supporting the findings of the study are available within the paper and its supplementary information files.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

B.E.B. declares outside interests in Fulcrum Therapeutics, 1CellBio, HiFiBio, Arsenal Biosciences, Cell Signaling Technologies, BioMillenia, and Nohla Therapeutics. H.M. is a paid employee of Daiichi Sankyo Co. The other authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

<sup>2</sup>Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>3</sup>Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

Received: 27 January 2020 Accepted: 25 June 2020

Published online: 15 July 2020

## References

- McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*. 2017;168(4):613–28.
- Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature*. 2017;543(7643):122–5.
- Nathanson DA, Gini B, Mottahedeh J, Visnyei K, Koga T, Gomez G, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*. 2014;343(6166):72–6.
- Salgia R, Kulkarni P. The genetic/non-genetic duality of drug 'resistance' in cancer. *Trends Cancer*. 2018;4(2):110–8.
- Oxnard GR. The cellular origins of drug resistance in cancer. *Nat Med*. 2016;22(3):232–4.
- Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010;141(1):69–80.
- Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. *Cell Stem Cell*. 2017;20(2):233–46. e7.
- Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883–92.
- Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90–4.
- Brocks D, Assenov Y, Minner S, Bogatyrova O, Simon R, Koop C, et al. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep*. 2014;8(3):798–806.
- Bhang HE, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med*. 2015;21(5):440–8.
- Almendo V, Cheng YK, Randles A, Itzkovitz S, Marusyk A, Ametller E, et al. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep*. 2014;6(3):514–27.
- Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol Cell*. 2014;54(5):716–27.
- Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, 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*. 2009;10(5):459–66.
- Stupp R, Taillibert S, Kanner A, Read W, Steinberg D, Lhermitte B, et al. Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA*. 2017;318(23):2306–16.
- Touat M, Idbaih A, Sanson M, Ligon KL. Glioblastoma targeted therapy: updated approaches from recent biological insights. *Ann Oncol*. 2017;28(7):1457–72.
- Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell*. 2014;14(3):275–91.
- Singh SK, Clarke ID, Terasaki M, Bonn VE, Hawkins C, Squire J, et al. Identification of a cancer stem cell in human brain tumors. *Cancer Res*. 2003;63(18):5821–8.
- Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A. Whole-organism clone tracing using single-cell sequencing. *Nature*. 2018;556(7699):108–12.
- Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol*. 2018;36(5):442–50.
- Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat Biotechnol*. 2018;36(5):469–73.
- Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*. 2018;360(6392):981–7.
- Schmitt MW, Loeb LA, Salk JJ. The influence of subclonal resistance mutations on targeted cancer therapy. *Nat Rev Clin Oncol*. 2016;13(6):335–47.
- Vanner RJ, Remke M, Gallo M, Selvadurai HJ, Coutinho F, Lee L, et al. Quiescent sox2(+) cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell*. 2014;26(1):33–47.
- Roesch A, Vultur A, Bogeski I, Wang H, Zimmermann KM, Speicher D, et al. Overcoming intrinsic multidrug resistance in melanoma by blocking the mitochondrial respiratory chain of slow-cycling JARID1B(high) cells. *Cancer Cell*. 2013;23(6):811–25.
- Chen J, Li Y, Yu TS, McKay RM, Burns DK, Kernie SG, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature*. 2012;488(7412):522–6.
- Gierahn TM, Wadsworth MH 2nd, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods*. 2017;14(4):395–8.
- Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010;463(7283):899–905.
- Brennan CW, Verhaak RG, McKenna A, Campos B, Nounshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155(2):462–77.
- Day E, Poullogiannis G, McCaughan F, Mulholland S, Arends MJ, Ibrahim AE, et al. IRS2 is a candidate driver oncogene on 13q34 in colorectal cancer. *Int J Exp Pathol*. 2013;94(3):203–11.
- Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and transcriptomic landscape of triple-negative breast cancers: subtypes and treatment strategies. *Cancer Cell*. 2019;35(3):428–40. e5.
- Liu TT, You HL, Weng SW, Wei YC, Eng HL, Huang WT. Recurrent amplification at 13q34 targets at CUL4A, IRS2, and TFDP1 as an independent adverse prognosticator in intrahepatic cholangiocarcinoma. *PLoS One*. 2015;10(12):e0145388.
- Melchor L, Saucedo-Cuevas LP, Munoz-Repeto I, Rodriguez-Pinilla SM, Honrado E, Campoverde A, et al. Comprehensive characterization of the DNA amplification at 13q34 in human breast cancer reveals TFDP1 and CUL4A as likely candidate target genes. *Breast Cancer Res*. 2009;11(6):R86.

34. Nishimura R, Takita J, Sato-Otsubo A, Kato M, Koh K, Hanada R, et al. Characterization of genetic lesions in rhabdomyosarcoma using a high-density single nucleotide polymorphism array. *Cancer Sci.* 2013;104(7):856–64.
35. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008;455(7216):1061–8.
36. Shaw CM, Grobmyer SR, Ucar DA, Cance WG, Reith JD, Hochwald SN. Elevated expression of IRS2 in the progression from neurofibroma to malignant peripheral nerve sheath tumor. *Anticancer Res.* 2012;32(2):439–43.
37. You HL, Liu TT, Weng SW, Chen CH, Wei YC, Eng HL, et al. Association of IRS2 overexpression with disease progression in intrahepatic cholangiocarcinoma. *Oncol Lett.* 2018;16(4):5505–11.
38. Szabolcs M, Keniry M, Simpson L, Reid LJ, Koujak S, Schiff SC, et al. Irs2 inactivation suppresses tumor progression in Pten+/- mice. *Am J Pathol.* 2009;174(1):276–86.
39. Dearth RK, Cui X, Kim HJ, Kuitatse I, Lawrence NA, Zhang X, et al. Mammary tumorigenesis and metastasis caused by overexpression of insulin receptor substrate 1 (IRS-1) or IRS-2. *Mol Cell Biol.* 2006;26(24):9302–14.
40. Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell.* 2006; 9(3):157–73.
41. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell.* 2017;32(1):42–56. e6.
42. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* 2010; 17(1):98–110.
43. Al-Salam A, Irwin DM. Evolution of the vertebrate insulin receptor substrate (Irs) gene family. *BMC Evol Biol.* 2017;17(1): 148.
44. Esposito DL, Aru F, Lattanzio R, Morgano A, Abbondanza M, Malekzadeh R, et al. The insulin receptor substrate 1 (IRS1) in intestinal epithelial differentiation and in colorectal cancer. *PLoS One.* 2012;7(4):e36190.
45. Dearth RK, Cui X, Kim HJ, Hadsell DL, Lee AV. Oncogenic transformation by the signaling adaptor proteins insulin receptor substrate (IRS)-1 and IRS-2. *Cell Cycle.* 2007;6(6):705–13.
46. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer.* 2019;19(5):283–8.
47. Galanis E, Anderson SK, Twohy EL, Carrero XW, Dixon JG, Tran DD, et al. A phase 1 and randomized, placebo-controlled phase 2 trial of bevacizumab plus dasatinib in patients with recurrent glioblastoma: Alliance/North Central Cancer Treatment Group N0872. *Cancer.* 2019;125(21):3790–800.
48. Koschmann C. Study of dasatinib in combination with everolimus for children and young adults with gliomas harboring PDGFR alterations [Available from: <https://clinicaltrials.gov/ct2/show/NCT03352427>.
49. Lassman AB, Pugh SL, Gilbert MR, Aldape KD, Geinoz S, Beumer JH, et al. Phase 2 trial of dasatinib in target-selected patients with recurrent glioblastoma (RTOG 0627). *Neuro-Oncology.* 2015;17(7):992–8.
50. Wakimoto H, Kesari S, Farrell CJ, Curry WT Jr, Zaupa C, Aghi M, et al. Human glioblastoma-derived cancer stem cells: establishment of invasive glioma models and treatment with oncolytic herpes simplex virus vectors. *Cancer Res.* 2009; 69(8):3472–81.
51. Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell.* 2016;167(7):1867–82. e21.
52. R Core Team. R: A language and environment for statistical computing. Vienna, Austria.
53. van Galen P, Hovestadt V, Wadsworth II MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-cell RNA-Seq reveals AML hierarchies relevant to disease progression and immunity. *Cell.* 2019;176(6):1265–81. e24.
54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
55. van der Maaten L. Accelerating t-SNE using tree-based algorithms. *J Mach Learn Res.* 2014;15:3221–45.
56. Krijthe JH. Rtsne: T-distributed stochastic neighbor embedding using Barnes-hut implementation; 2015.
57. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature.* 2016;539(7628):309–13.
58. Creighton CJ, Fu X, Hennessy BT, Casa AJ, Zhang Y, Gonzalez-Angulo AM, et al. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res.* 2010;12(3):R40.
59. Fan J. In: Yuan G, editor. Computational methods for single-cell data analysis; 2019.
60. Fan J. LIGER: Lightweight Iterative Geneset Enrichment. 1.1 ed; 2017.
61. van Galen P, Kreso A, Mbong N, Kent DG, Fitzmaurice T, Chambers JE, et al. The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature.* 2014;510(7504):268–72.
62. Eyler CE, Matsunaga H, Hovestadt V, Vantine SJ, van Galen P, Bernstein BE. Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance. Datasets. Gene Expression Omnibus. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142119> (2019).

## Publisher's Note

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