Extrachromosal DNA elements can drive disease evolution in glioblastoma

Ana C. deCarvalho[1*†], Hoon Kim[2*], Laila M. Poisson[3], Mary E. Winn[4], Claudius Mueller[5], David Cherba[4], Julie Koeman[6], Sahil Seth[7], Alexei Protopopov[7], Michelle Felicella[8], Siyuan Zheng[2], Jianhua Zhang[7], Emanuel F. Petricoin[5], Lynda Chin[2, 7], Tom Mikkelsen[1, 9†], Roel G.W. Verhaak[2, 10†]

[1]Department of Neurosurgery, Henry Ford Hospital, Detroit, MI 48202, USA.
[2]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston , TX 77030, USA.
[3]Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI 48202, USA.
[4]Bioinformatics and Biostatistics Core, Van Andel Research Institute, Grand Rapids, MI 49503, USA.
[5]Center for Applied Proteomics and Personalized Medicine, George Mason University, Manassas, VA, USA
[6]Pathology and Biorepository Core, Van Andel Research Institute, Grand Rapids, MI 49503, USA.
[7]Institute for Applied Cancer Science, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.
[8]Department of Pathology, Henry Ford Hospital, Detroit, MI 48202, USA.
[9]Department of Neurology, Henry Ford Hospital, Detroit, MI 48202, USA.
[10]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston , TX 77030, USA.

* These authors contributed equally to this work.
† Correspondence: adecarv1@hfhs.org (A.C.D.), tmikkel1@hfhs.org (T.M.), rverhaak@mdanderson.org (R.G.W.V.)

Keywords: glioblastoma, double minute, tumor evolution, MET

ABSTRACT

To understand how genomic heterogeneity of glioblastoma (GBM) contributes to the poor response to therapy which is characteristic of this disease, we performed DNA and RNA sequencing on primary GBM, neurospheres and orthotopic xenograft models derived from the same parental tumor. We used these data to show that somatic driver alterations were in majority propagated from tumor to model systems. In contrast, we found that amplifications of *MET*, a proto-oncogene coding for a receptor tyrosine kinase, were detected in three of thirteen primary GBM, largely discarded in neurospheres cultures, but resurfaced in xenografts . We inferred the clonal evolution dynamics of all models using somatic single nucleotide variants (sSNVs) and were unable to find sSNVs replicating the pattern delineated by the *MET* amplification event despite its strong selective effect. FISH analysis showed that most copies of *MET* resided on extrachromosomal DNA elements commonly referred to as double minutes. Long range Pacific Biosciences sequencing recovered the *MET* containing circular double minute structure. The evolutionary propagation patterns suggested that *MET* double minutes and sSNVs were disjointly inherited. The context-dependent re-emergence of *MET* amplifiations suggests that the microenvironmental milieu was a critical regulator of extrachromosomal *MET* driven cell proliferation. Our analysis shows that extrachromosomal elements are able to drive tumor evolution.

INTRODUCTION

Cancer genomes are subject to continuous mutagenic processes in combination with an inability to repair DNA damage [1]. Somatic genomic variants that are acquired throughout tumorigenesis may provide cells with a competitive advantage over their neighboring cells in the context of a nutrition- and oxygen-poor microenvironment, resulting in increased survival and/or proliferation rate [2]. The Darwinian evolutationary process results in intratumoral heterogeneity in which single cancer cell derived tumor subclones are characterized by unique somatic alterations [3]. Chemotherapy and ionizing radiation may enhance intratumoral evolution by eliminating cells lacking the ability to deal with increased levels of genotoxic stress, while targeted therapy may favor subclones in which the targeted vulnerability is absent [4,5]. Increased clonal heterogeneity has been associated with tumor progression and mortality [6]. Computational methods that analyze the allelic fraction of somatic variants identified from high throughput sequencing data sets are able to infer clonal population structures and provide insights into the level of intratumoral clonal variance [7].

Glioblastoma (GBM), a WHO grade IV astrocytoma, is the most prevalent and aggressive primary central nervous system tumor. GBM is characterized by poor response to standard post-resection radiation and cytotoxic therapy, resulting in dismal prognosis with a 2 year survival rate around 15% [8]. The genomic and transcriptomic landscape of GBM has been extensively described [9-11]. Intratumoral heterogeneity in GBM has been well characterized, in particular with respect to somatic alterations affecting receptor tyrosine kinases [12-14]. To evaluate how genomically heterogeneous tumor cell populations are affected by selective pressures arising from the transitions from tumor to culture to xenograft, we performed a comprehensive genomic and transcriptomic analysis of thirteen GBMs, the glioma-neurosphere forming cultures (GSC) derived from them, and xenograft models (PDX) established from early passage neurospheres. Our results highlight the evolutionary process of GBM cells as they are propagated from primary tumor to *in vitro* neurosphere culture and transplanted to *in vivo* models, placing great emphasis on the role of extrachromosomal elements in driving tumor formation which establishes these cellular bodies as potential targets for new cancer therapeutics.

RESULTS

Genomic profiling of glioblastoma, derived neurosphere and PDX samples

We established neurosphere cultures from 12 newly diagnosed and one matched recurrent GBM (Table 1). Neurosphere cultures between 7 and 18 passages were used for molecular profiling and engrafting orthotopically into nude mice. The sample cohort included one pair of primary (HF3016) and matching recurrent (HF3177) GBM. A schematic overview of our study is presented in Figure 1.

Low pass sequencing at a median depth of 6.5X (+/- 1.8) was performed and was used to determine the genome wide DNA copy number profile of all samples. DNA copy number was generally highly preserved between tumor and derived model systems (Fig. 2a). Whole chromosome 7 gain and chromosome 10 loss were retained in model systems when detected in the tumor, consistent with their proposed role as canonical GBM lesions that occur amongst the earliest events in gliomagenesis [15]. Analysis of B-allele fractions revealed loss of heterozygosity (LOH) of chromosome 10 in two cases with diploid chromosome 10, suggesting these cases had first lost a single copy of the chromosome which was subsequently duplicated (Extended Data Fig. 1a). We evaluated chromosome 10 LOH using Affymetrix SNP6 profiles from 320 IDH-wildtype TCGA glioblastoma, and found that 27 of 52 tumors with diploid chromosome 10 similarly showed LOH, underscoring the importance of aberrations in chromosome 10 in gliomagenesis and evolution (Extended Data Fig. 1b). The global DNA copy number resemblance between xenografts and the GBM from which they were derived confirms that PDXs recapitulate the majority of molecular properties found in the original tumor.

To determine whether model systems capture the genes that are thought to drive gliomagenesis, and whether there is selection for specific driver genes, we performed exome sequencing on all samples and jointly compared mutation and DNA copy number status of genes previously found to be significantly mutated, gained, or lost in GBM [9,11]. We found that 100% of homozygous deletions and somatic single nucleotide variants (sSNVs) affecting GBM driver genes in tumor samples were propagated to the neurospheres and xenografts, including non-coding variants in the *TERT* promoter (Fig.

2b). Genomic amplifications showed greater heterogeneity. *MYC* amplifications were acquired in three neurospheres and maintained in xenografts in two out of three, consistent with its role in glioma stem cell maintenance [16,17]. Other genes showing variable representation across tumor and model systems included *MET*, *EGFR* and *PIK3CA*.

A single primary GBM, HF2354, was subjected to neoadjuvant carmustine treatment and its derived model systems were considerably less similar compared to the primary tumor than other cases. Whole chromosome gains of chromosome 1, 14 and 21, and one copy loss of chromosome 3, 8, 13, 15 and 18 were acquired in the neurosphere culture and propagated to the xenograft models. At the gene level, this resulted in newly detected mutations in *PTEN* and *TP53*, focal amplification of *MYC*, and absence of *CDK4* and *EGFR* amplification in the neurosphere and xenografts relative to the tumor sample.

Extrachromosomal CAPZA2-MET fusion transcripts regulate *in vivo* tumor evolution

Chimeric RNA fusions have been previously reported in GBM [18-20] and may be therapeutically targetable, in particular when involving receptor tyrosine kinases [21,22]. We performed RNA sequencing and detected fusion transcripts in all samples except for a single neurosphere line (HF3203) with disqualifying quality control values [23]. From this unbiased screen, multiple fusions joining the *CAPZA2* coding start with the 5' UTR of *MET* were identified in the primary tumors of HF3035, HF3077 and HF3055 (Fig. 3a). Additional *CAPZA2-MET* variants resulted in an in-frame transcript consisting of *CAPZA2* exon 1 and *MET* starting from exon 3 (HF3035, HF3077) and exon 6 (HF3035). The *CAPZA2-MET* fusions associated with outlier gene expression of *MET*, suggestive that the fusion resulted in MET activation (Extended Data Fig. 2a). *CAPZA2* expression was comparable between samples with and without *CAPZA2-MET* fusions. The presence of multiple parallel fusion transcripts suggested complex chromosomal rearrangements, which associated with focal amplification of a 200 kb area on 7q31 (Fig. 3a). Co-amplification of the 7q31 genomic area carrying the adjacent *CAPZA2* and *MET* genes has been previously reported in glioma [24]. To assess the frequency of MET-

activating somatic alterations in glioblastoma we analyzed the DNA copy number profiles of 486 TCGA IDH wildtype glioblastoma samples. A focal amplification of the *MET* locus ranging in size from 150kb to 5.1 Mb which associated with a highly significant increase in expression relative to samples with broad 7q amplification or diploid *MET* copy number was identified in ten cases (2.1%) (Extended Data Fig. 2b). RNAsequencing data was available for one of the ten TCGA cases and no fusions involving *MET* were detected in that sample. *CAPZA2-MET* fusions have been infrequently reported in other cancers [25,26]. Clinical response of a glioblastoma carrying *MET* amplification in response to MET and ALK inhibiting agent crizotinib has been recorded [27].

In spite of convincing evidence supporting fusion events in the GBM samples from HF3035, HF3055 and HF3077, no sequencing reads manifesting the presence of *CAPZA2-MET* fusions were identified in the HF3055 and HF3077 neurospheres and only weak support was found in the HF3035 neurosphere. However, identical *CAPZA2-MET* fusions resurfaced at high frequency in all xenografts derived from the HF3035 and HF3077 neurospheres. None of the HF3055 xenografts carried *CAPZA2-MET* fusions, in line with the absence of focal *MET* amplification in the primary HF3055 tumor. To exclude the possibility that the *CAPZA2-MET* fusion events were artifacts resulting from sequencing we validated the event in all samples from HF3035 using RT-PCR, which confirmed both wildtype MET and *CAPZA2-MET* mRNA in the tumor and PDX but not neurosphere (Extended Data Fig. 2c). The DNA breakpoints resulting in a *CAPZA2-MET* were found to be different between HF3035, HF3077, and HF3055 confirming that these were independent events and not the result of sample contamination. MET protein was abundantly present in the HF3035 and HF3077 tumors as measured using immunohistochemistry, undetectable in the neurospheres and re-expressed in the PDX (Extended Data Fig. 2d).

The pattern of disappearing and re-appearing MET rearrangements may result from clonal selection of glioblastoma cells with a competitive advantage for proliferation *in vivo*. This hypothesis is strengthened by the observation that MET is a growth factor responsive cell surface receptor tyrosine kinase [28]. We reasoned that evolutionary patterns resulting in such dominant clonal selection would likely be replicated by sSNVs

tracing the cells carrying the MET amplicon. To evaluate clonal selection patterns, we determined variant allele fractions of all sSNVs identified across in HF3035 and HF3077 samples. To increase our sensitivity to detect mutations present in small numbers of cells, we corroborated the exome sequencing data using high coverage (>1,400x) targeted sequencing. All mutations detected in the HF3035 GBM were recovered in the neurosphere and xenografts. The mutational profile of HF3035 suggested that a subclone developed in the xenografts that was not present in parental GBM and neurosphere and revealed a subclone that was present at similar frequencies in all samples. Only a single and very low frequency *LAMB1* mutation (variant allele fraction = 0.003) present in the HF3077 primary tumor, but not detected in its derived neurosphere, resurfaced in one of three xenografts with a 0.04 variant allele fraction. A low frequency subclone (C2) developed in the neurosphere which was transmitted to xenografts (Fig. 3b). Subclonal heterogeneity as recovered by the mutation profiles thus suggested a very different clonal selection trend compared to to the disappearing and resurfacing *MET* amplifications and associated gene fusions.

We hypothesized that the discordance of evolutionary pattern between *CAPZA2-MET* fusions and somatic point mutations could be explained by the presence of extrachromosomal DNA segments known as double minutes. Double minutes are thought to originate from genomic amplifications through post-replicative excision of chromosomal fragments and non-homologous end joining [29]. They typically exist as episomes which are inherited through random distribution over the two daughter cells[30], possible through a binomial model[31]. Double minutes have been identified in 10-15% of glioblastoma [19,32], with evidence indicating they may get depleted in culture [33] but retained in PDX models [29]. The presence of double minutes has been associated with worse outcome, copy number instability, and increased mutational rates [19,33]. We employed fluorescence *in situ* hybridization (FISH) using probes targeting MET and included probes targeting the chromosome 7 centromere for comparison. Interphase FISH analysis confirmed that *MET* amplification was present in a much higher number of cells from primary GBM and PDX models compared to neurosphere in HF3035 and HF3077 (Fig. 3c; Supplementary Table 1). To determine whether *MET* amplification occurred extrachromosomally, we generated metaphase spreads using live cells

obtained from early passage neurosphere cultures established from the original tumor (P5 and P7) and from two HF3035 PDX tumors (PDX_NS1 and PDX_NS2). We analyzed metaphase-enriched nuclei and observed that *MET* DNA is frequently extrachromosomal (Fig. 3c). A small percentage of nuclei with 3 copies of chromosome 7 but only only 2 copies of MET was detected in HF3077 tumor. The frequency of cells with one deleted copy of MET increased significantly in HF3077 neurospheres and was also observed in the xenografts.

To precisely define the genomic contents and structure of the predicted double minutes, we generated long read (Pacific Biosciences) DNA sequencing from a single HF3035 and HF3077 xenograft, and performed *de novo* assembly. In HF3035, seven assembled contigs (range: 6,466 ~ 135,621 bp) were identified to have sequence fragments (at least 1,000 bp long) aligned on the *MET-CAPZA2* region of hg19 chromosome 7. Interestingly, analysis of the aligned sequence fragments from the seven contigs revealed a very complex structural rearrangement than expected from the analysis of short read sequencing data (Extended Data Fig. 3a). For example, the 135kb tig01170337 contig consisted of 8 sequence framents that were nonlinearly aligned on alternating strands of the *MET-CAPZA2* and *CNTNAP2* regions. Other contigs such as tig01170699, tig01170325, and tig00000023 also showed nonlinear alignment, suggesting that these contigs resulted from associated and potentially chromosomal structural variations. We performed pairwise sequence comparison of the contigs to search for sequence fragments (at least 5,000 bp long) shared among them, and we found four contigs each of which shared sequence fragments with one of the contigs. Interestingly, three of them could be connected in a circular form using the shared sequence fragments (Fig. 3d), revealing a circular structure that may represent the full double minute. In HF3077, only two contigs were detected to be aligned on the *MET-CAPZA2* region of hg19 chromosome 7 (Fig. 3d; Extended Data Fig. 3a). Presence of only two aligned contigs in HF3077 might be related to the lower sequence coverage of the double minute structure, compared to HF3035 (34x vs 405x, respectively) (Extended Data Fig. 3b). The longest contig, tig01141776 (183,455 bp long), consisted of two segment framents that were nonlinearly aligned over exon 1 of *CAPZA2* and all except exons 3-5 of *MET*, suggesting that it resulted from structural

variations. The second short contig, tig01141835 (22,628 bp long), was aligned as a whole over exon 3-5 of *MET*. Interestingly, connecting the two contigs created a circular DNA segment. Through analysis of PacBio sequencing, we were able to detect and reconstruct the predicted double minute structures.

Extrachromosomal amplification of Abelson murine leukemia viral oncogene homolog 1 (ABL1) resulting in *NUP214-ABL1* transcript fusions have been described in T-cell acute lymphoblastic leukemias [34]. Regulation of extrachromosomal *EGFR* vIII variants associated with resistance mechanisms to EGFR inhibitors such as lapatanib were reported in glioblastoma samples [35]. Our data indicates that dynamic regulation of MET carrying extrachromosomal DNA represents a driver of clonal selection with microenvironmental exposures providing a critical mediator of extrachromosomal MET driven cell proliferation. The discordance in propagation patterns between sSNVs and *MET* carrying extrachromosomal elements suggested that they were inherited disjointly. Where sSNVs are copied to daughter cells in a manner reminiscent of the Mendelian rules described for alleles segregating during meiosis such that each gamete only carries a single allele, MET DMs likely randomly segregated.

Large, megabase sized double minutes are frequently found in glioblastoma and can be identified using whole genome sequencing and DNA copy number data [19,32,33]. To determine whether double minutes can survive therapeutical barriers and thereby drive clonal selection, we evaluated the DNA copy number profiles of 23 matching pairs of primary and recurrent GBM for the presence of double minutes [4]. Evidence supporting the presence of double minutes was found in five of 23 primary tumors. The DNA copy number alteration pattern suggestive of a double minute was preserved after disease recurrence in three of five pairs (Extended Data Fig. 4), supporting the notion that double minutes can prevail after the selective pressures imposed by therapy. One of 23 primary GBMs carried a focal *MET* amplification which associated with a *NRCAM-MET* transcript fusion. *NRCAM* resides 8Mb downstream of *MET* on chromosome 7. A lesion with similar DNA breakpoints and leading to overexpression of *MET* was detected in the matching recurrence suggesting that MET was a driver of gliomagenesis in this case. We were unable to determine whether this MET lesion was episomal on the basis of the DNA copy number profile and in absence of WGS data for this pair.

Clonal drift from parental tumor to neurosphere culture

The precise genomic characterization of our sample cohort allowed us to evaluate evolutionary distance from parental tumors to their freshly derived tumor models. The average mutation frequency (measured as number of sSNVs per Mb) per parent GBM was 0.44 (range: 0.42~0.74) which was elevated to 0.54 sSNVs per Mb (range : 0.31~0.96) in neurospheres and 0.52 sSNVs per Mb (range : 0.31~1.32) in xenografts (Fig. 2b; Supplementary Table 2). Approximately 80% (range: 60%~95%) of mutations detected in parent tumors were retained in neurospheres, demonstrating genetic proximity between patient tumors and their neurospheres. In order to determine whether sSNVs in xenografts and neurospheres were newly acquired after culturing/transplantation or existed at frequencies below the detection threshold of exome sequencing, we performed deep coverage targeted sequencing (>1,400x) on 792 mutations distributed over the thirteen sample sets. Validation sequencing recovered 21% of exome sequencing mutations detected in the parent tumor that were not called in the neurosphere, showing that culturing of GBM cells under serum free conditions results in some loss of genomic characteristics. Only 7% of exome sequencing mutations detected in neurosphere but not parent tumor were recovered by validation sequencing in the parent tumor, suggesting that these were in majority sSNVs acquired during cell culture. To compare the level of intratumoral heterogeneity between samples, we clustered sSNV tumor cell fractions inferred from variant allele frequencies (Fig. 4a; Extended Data Fig. 5) [7]. Only two of thirteen sample sets (HF2587 and HF3178) were found to lose a cluster of mutations at similar tumor cell fractions, suggesting that a subclonal cell population was not transferred between parent GBM and neurosphere. Five of thirteen neurospheres had acquired a new subclone. Similar levels of heterogeneity have been reported in multisector sequencing of GBM samples [4], suggesting the similarity between parental tumor and neurosphere culture may also be explained by variability between the tumor portion used for exome sequencing and the part of the tumor used to establish the neurosphere culture.

Heterogeneous xenograft tumors derived from a single neurosphere

Three xenografts from each neurosphere model were analyzed. Subclonal mutation clusters were detected in six neurospheres and were propagated to multiple xenografts in five of six cases. On average 80% (range: 33% to 100%) of neurosphere mutations were transmitted to PDX models, and propagated mutations were stably detected in all three biological replicates. The fraction of parental GBM mutations recovered in PDXs ranged from 53% to 98%, average 81%, which is comparable to the fraction of mutations that was retained in the neurospheres. Xenograft models acquired on average 18 mutations compared to neurospheres, an increased mutation frequency of 34%, and representing an active mutational process during in vivo growth [36]. Amongst the acquired mutations were only a few genes marked as glioma driving [9]. An example of convergent evolution was seen in HF3160, where variable deletions of *CDKN2A*, encoding for p16, were detected in neurosphere and PDXs. However, in most cases we only noted limited clonal dynamics between neurosphere and PDX models.

Included in our collection were a pair of primary and recurrent GBM, respectively HF3016 and HF3177. This set provided an opportunity to compare natural disease evolution during chemo- and radio-therapy versus the artificial evolution establishing in the respective tumor models. While primary and recurrent tumor were globally very similar (Fig. 2b), a focal *MYC* amplification not detected in the HF3016 tumor was present in the neurosphere and PDXs derived from this tumor, and was also detected in all samples from the recurrent tumor. Enrichment of cells with *MYC* amplification in neurosphere cultures form GBM tumors has been observed by others [37], consistent with somatic alterations associated with an aggressive phenotype driving proliferation of both model systems as well as disease recurrence, suggesting a role for the development of tumor models in clinical practice.

Discussion

Glioblastoma is a heterogeneous disease that is highly resistant to chemo- and radiotherapy. New modalities for treatment are urgently needed. Modeling of tumors through cell culture and orthotopic xenotransplantation are essential approaches for preclinical target screening and validation, but in GBM have yet to result in novel therapies. Whether these models truthfully recapitulate the parental tumor is a topic of active discussion. Here, we showed that neurosphere models and intracranial xenografts are genomically similar, capturing over 80% of genomic alterations detected in parental tumors.

A surprising observation in our study was the divergence in propagation of structural rearrangements and point mutations, particularly focused on a focal MET amplification identified in three out of 13 GBMs. While this *MET* lesion was not or weakly detected in the derived neurospheres, an identical *MET* event was found in six of nine PDX models. This pattern is strongly suggestive of environmental pressures that favor *MET*-wildtype cells in culture but *MET*-activated cells in vivo. FISH analysis suggested that the *MET* gene was amplified through an extrachromosomal double minute structure which was reconstructed in two representative xenografts using long read sequencing. Mitosis results in equal representation of all chromosomal genotypes between daughter cells. However, extrachromosomal elements randomly distribute over daughter cells. This may explain why the evolution of the *MET* event was not similarly captured by sSNVs detected in parental tumor and PDX models but not neurospheres (Fig. 5). This finding has two important implications: a. it shows that the tumor microenvironment can play a critical role in clonal selection and suggests that targeting pro-tumor stimuli from the microenvironment is a viable treatment strategy; and b. tumor evolution can be dominated by extrachromosomal elements, showing that detection of point mutations alone is insufficient to accurately delineate this process. Previous studies have found that extrachromosomal bodies can provide a reservoir for therapeutically targetable genomic alterations [35]. Targeted MET inhibition of *MET* amplified GBMs has shown clinical promise [27]. We propose that double minutes and episomes are capable of driving gliomagenesis. Breaking the mechanism by which

extrachromosomal elements get released from their parental chromosomes may represent a relevant treatment option.

This is the first reported re-emergence of a double minute oncogene amplification upon orthotopic transplantation of GSCs with dramatically reduced frequency of nuclei carrying amplified *MET*. While the analysis described here has focused on *MET*, double minutes have been reported in 10-15% of GBM [19,32,33]. These lesions most frequently involved genes on chromosome 12p, including *CDK4* and *MDM2*, span up to several megabases in size, and can be recognizing by an intermittent amplification-deletion DNA copy number pattern. An important characteristic of our finding is the size of the *MET* double minute. We were unable to relate the *MET* event to other areas on the genome suggesting that this double minute is small and highly focal and exists in absence of the typical intermittent amplification-deletion pattern. Kb-sized single segment episomes can only be identified using high throughput sequencing approaches and therefore have been less frequently reported. Whether double minute size affects the mechanism of gliomagenesis is unclear. Extrachromosomal DNA is an understudied domain in cancer. Our analysis emphasizes the importance of this genomic alteration category for gliomagenesis. Future studies that specifically target the formation of episomal events may lead to therapies to prevent this process from happening. The models we described here may play a pivotal role in evaluating the potential of such approaches.

Acknowledgments

The authors would like to thank Dr. Norman Lehman and Dr. Chunhai (Charlie) Hao for pathology reviews; Lisa Scarpace for clinical information; Susan Irtenkauf, Laura Hasselbach, Kevin Nelson, Kimberly Bergman, and Susan Sobiechowski for cell culture and animal work; Andrea Transou, Yuling Meng, and Enoch Carlton for histology at HFH. We thank Genevieve Geneau, Sharen Roland, and Pac Bio platform personnel of the Génome Québec/Genome Canada-funded Innovation Centre for providing Pacific Biosciences sequencing. This work was supported by the LIGHT Research Program at the Hermelin Brain Tumor Center; grants from the National Institutes of Health P50 CA127001, R01 CA190121, and P01 CA085878; the Cancer Prevention & Research

Institute of Texas (CPRIT) R140606. We are hugely indebted to the patients who provided tumor and germline material for the purpose of this study.

Author Information

BAM files from exome sequencing, low pass whole genome sequencing and RNA sequencing used in this study were deposited to the European Genome-phenome Archive (EGA; http://www.ebi.ac.uk/ega/), which is hosted by the EBI and the CRG, under accession number EGAS00001001878. The authors declare no competing financial interests.

Supplementary Information

Extended Data Figures, Methods, and Supplementary Tables are available as supplementary data.

References

1 Roos, W. P., Thomas, A. D. & Kaina, B. DNA damage and the balance between survival and death in cancer biology. *Nat Rev Cancer* **16**, 20-33, doi:10.1038/nrc.2015.2 (2016).

2 Yap, T. A., Gerlinger, M., Futreal, P. A., Pusztai, L. & Swanton, C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med* **4**, 127ps110, doi:10.1126/scitranslmed.3003854 (2012).

3 Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N Engl J Med* **368**, 842-851, doi:10.1056/NEJMra1204892 (2013).

4 Kim, H. *et al.* Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res* **25**, 316-327, doi:10.1101/gr.180612.114 (2015).

5 Sequist, L. V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med* **3**, 75ra26, doi:10.1126/scitranslmed.3002003 (2011).

6 Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med* **22**, 105-113, doi:10.1038/nm.3984 (2016).

7 Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* **11**, 396-398, doi:10.1038/nmeth.2883 (2014).

8 Dolecek, T. A., Propp, J. M., Stroup, N. E. & Kruchko, C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005-2009. *Neuro Oncol* **14 Suppl 5**, v1-49, doi:10.1093/neuonc/nos218 (2012).

9 Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550-563, doi:10.1016/j.cell.2015.12.028 (2016).

10 Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98-110, doi:10.1016/j.ccr.2009.12.020 (2010).

11 Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462-477, doi:10.1016/j.cell.2013.09.034 (2013).

12 Snuderl, M. *et al.* Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**, 810-817, doi:10.1016/j.ccr.2011.11.005 (2011).

13 Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014, doi:10.1073/pnas.1219747110 (2013).

14 Szerlip, N. J. *et al.* Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc Natl Acad Sci U S A* **109**, 3041-3046, doi:10.1073/pnas.1114033109 (2012).

15 Ozawa, T. *et al.* Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. *Cancer Cell* **26**, 288-300, doi:10.1016/j.ccr.2014.06.005 (2014).

16 Wang, J. *et al.* c-Myc is required for maintenance of glioma cancer stem cells. *PloS one* **3**, e3769, doi:10.1371/journal.pone.0003769 (2008).

17 Annibali, D. *et al.* Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis. *Nat Commun* **5**, 4632, doi:10.1038/ncomms5632 (2014).

18 Singh, D. *et al.* Transforming fusions of FGFR and TACC genes in human glioblastoma. *Science* **337**, 1231-1235, doi:10.1126/science.1220834 (2012).

19 Zheng, S. *et al.* A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. *Genes Dev* **27**, 1462-1472, doi:10.1101/gad.213686.113 (2013).

20 Bao, Z. S. *et al.* RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res* **24**, 1765-1773, doi:10.1101/gr.165126.113 (2014).

21 Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, doi:10.1038/onc.2014.406 (2014).

22 Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* **15**, 371-381, doi:10.1038/nrc3947 (2015).

23 Torres-Garcia, W. *et al.* PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics* **30**, 2224-2226, doi:10.1093/bioinformatics/btu169 (2014).

24 Mueller, H. W. *et al.* Identification of an amplified gene cluster in glioma including two novel amplified genes isolated by exon trapping. *Hum Genet* **101**, 190-197 (1997).

25 Kim, H. P. *et al.* Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene* **33**, 5434-5441, doi:10.1038/onc.2013.490 (2014).

26 Yoshihara, K. *et al.* The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* **34**, 4845-4854, doi:10.1038/onc.2014.406 (2015).

27 Chi, A. S. *et al.* Rapid radiographic and clinical improvement after treatment of a MET-amplified recurrent glioblastoma with a mesenchymal-epithelial transition inhibitor. *J Clin Oncol* **30**, e30-33, doi:10.1200/JCO.2011.38.4586 (2012).

28 Organ, S. L. & Tsao, M. S. An overview of the c-MET signaling pathway. *Ther Adv Med Oncol* **3**, S7-S19, doi:10.1177/1758834011422556 (2011).

29 Vogt, N. *et al.* Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. *Proc Natl Acad Sci U S A* **101**, 11368-11373, doi:10.1073/pnas.0402979101 (2004).

30 Storlazzi, C. T. *et al.* Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Res* **20**, 1198-1206, doi:10.1101/gr.106252.110 (2010).

31 Lundberg, G. *et al.* Binomial mitotic segregation of MYCN-carrying double minutes in neuroblastoma illustrates the role of randomness in oncogene amplification. *PloS one* **3**, e3099, doi:10.1371/journal.pone.0003099 (2008).

32 Sanborn, J. Z. *et al.* Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res* **73**, 6036-6045, doi:10.1158/0008-5472.CAN-13-0186 (2013).

33 Nikolaev, S. *et al.* Extrachromosomal driver mutations in glioblastoma and low-grade glioma. *Nat Commun* **5**, 5690, doi:10.1038/ncomms6690 (2014).

34 Graux, C. *et al.* Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia. *Nat Genet* **36**, 1084-1089, doi:10.1038/ng1425 (2004).

35 Nathanson, D. A. *et al.* Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72-76, doi:10.1126/science.1241328 (2014).

36 Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422-426, doi:10.1038/nature13952 (2015).

37 Wakimoto, H. *et al.* Maintenance of primary tumor phenotype and genotype in glioblastoma stem cells. *Neuro-oncology* **14**, 132-144, doi:10.1093/neuonc/nor195 (2012).

Table 1. Clinical characteristics of GBM patients included in this study.

Sample	Pathology	Age/ Gender	Rx prior to surgery	MGMT	OS (days)	TTP (days)
HF2354	GBM	61/M	BCNU	U	196	60
HF2587	GBM	56/F	untreated	M	360	232
HF2927	GBM	55/F	untreated	U	664	566
HF3016	GBM	45/M	untreated	U	649	88
HF3177	rGBM4		RT/TMZ/DCVax	U		
HF3035	GBM	54/F	untreated	U	352	196
HF3055	GBM	58/M	untreated	U	371	77
HF3077	GBM	56/F	untreated	U	465	54
HF3160	GBM	21/F	untreated	M	1018	100
HF3178	GBM	65/M	untreated	U	189	138
HF3203	GBM	64/M	untreated	U	425	276
HF3216	GBM	76/M	untreated	U	94	
HF3253	GBM	82/F	untreated	U	68	

Rx: treatment; MGMT: *MGMT* gene methylation status, U = unmethylated, M = methylated; OS: overall survival; TTP: time to progression.

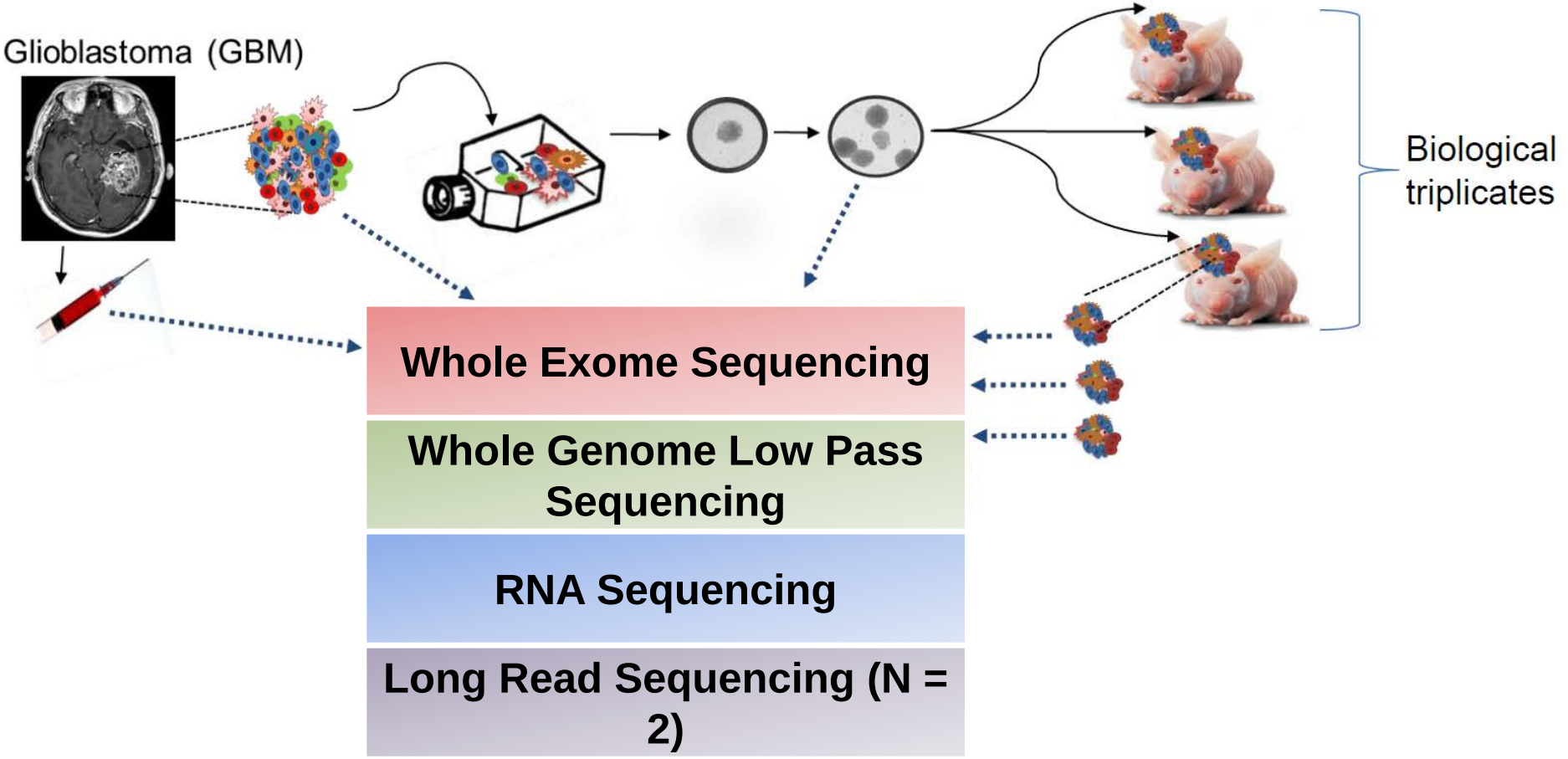


Figure 1. Schematic study overview.

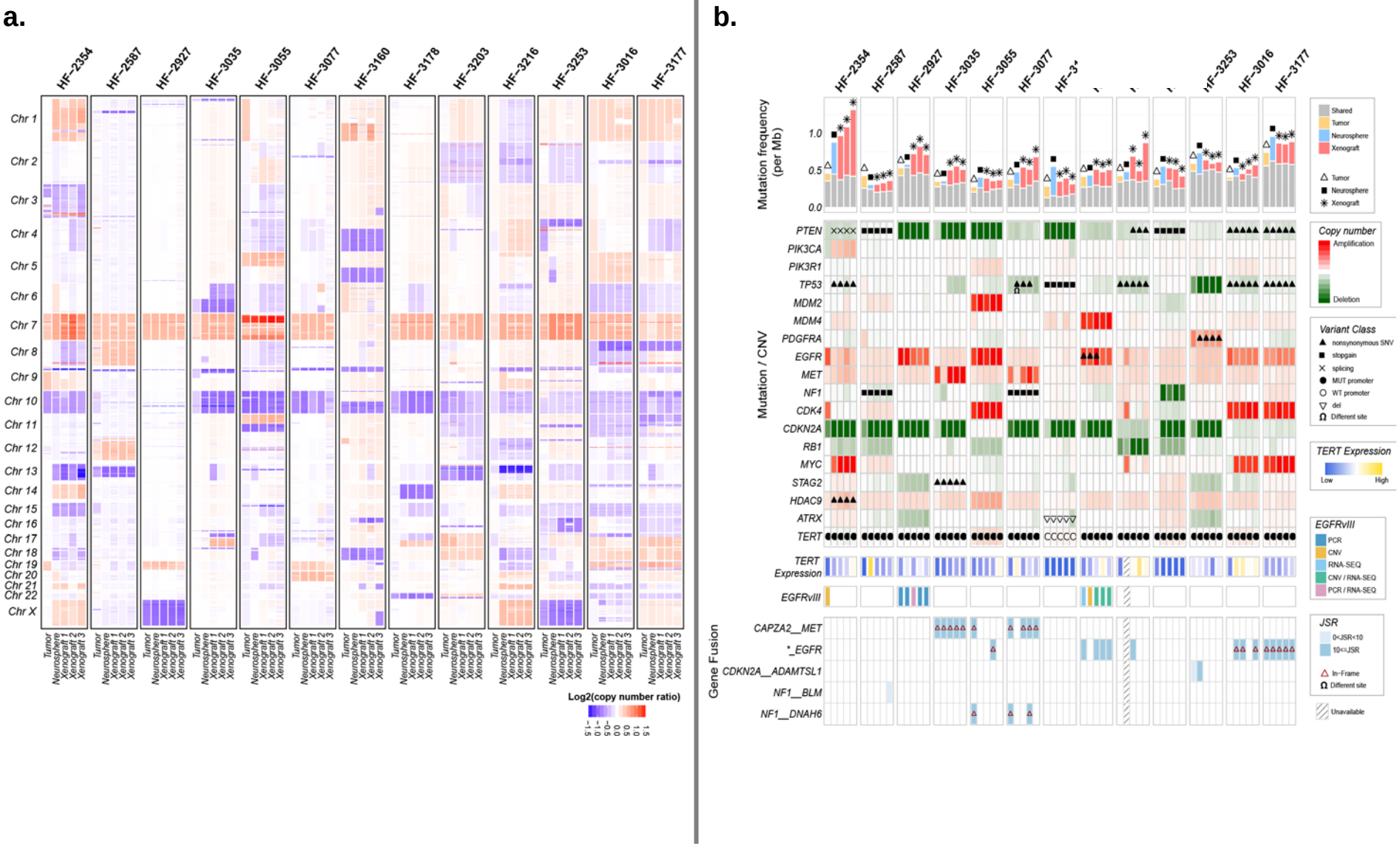


Figure 2. Somatic alteration landscape of GBM and model systems. a. Genome wide DNA copy number profiles. Copy number increases (red) and decreases (blue) are plotted as a function of distance along the normal genome (vertical axis, divided into chromosomes). **b.** Somatic driver alterations compared between GBM tumors and derived model systems.

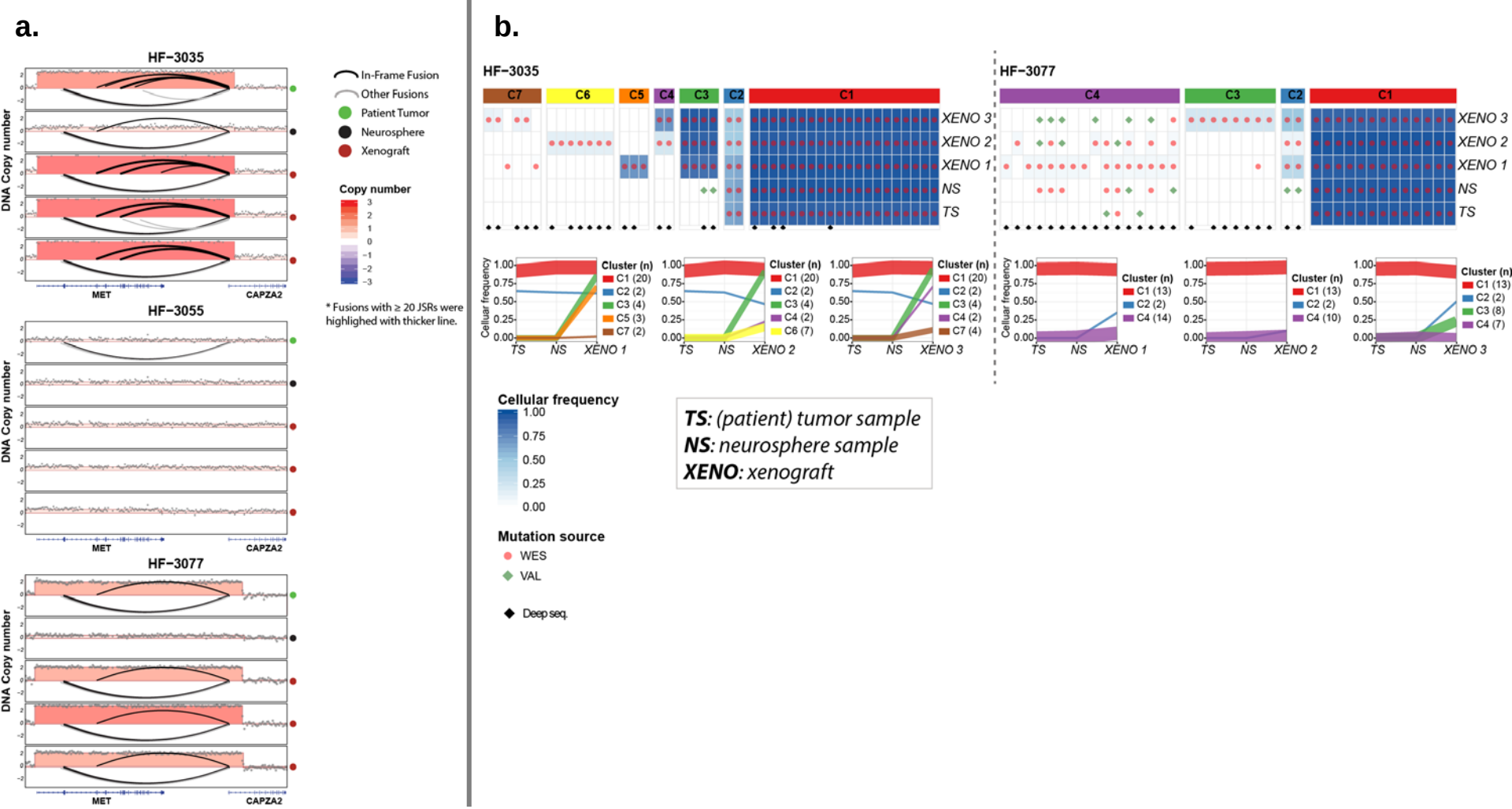
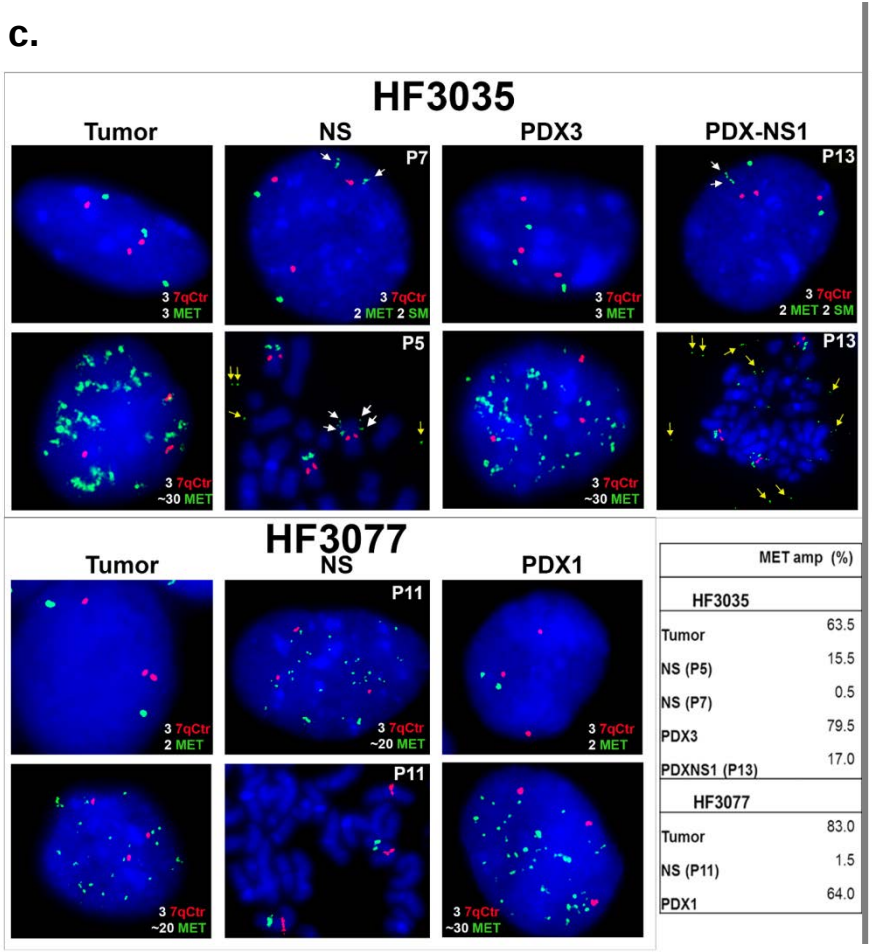
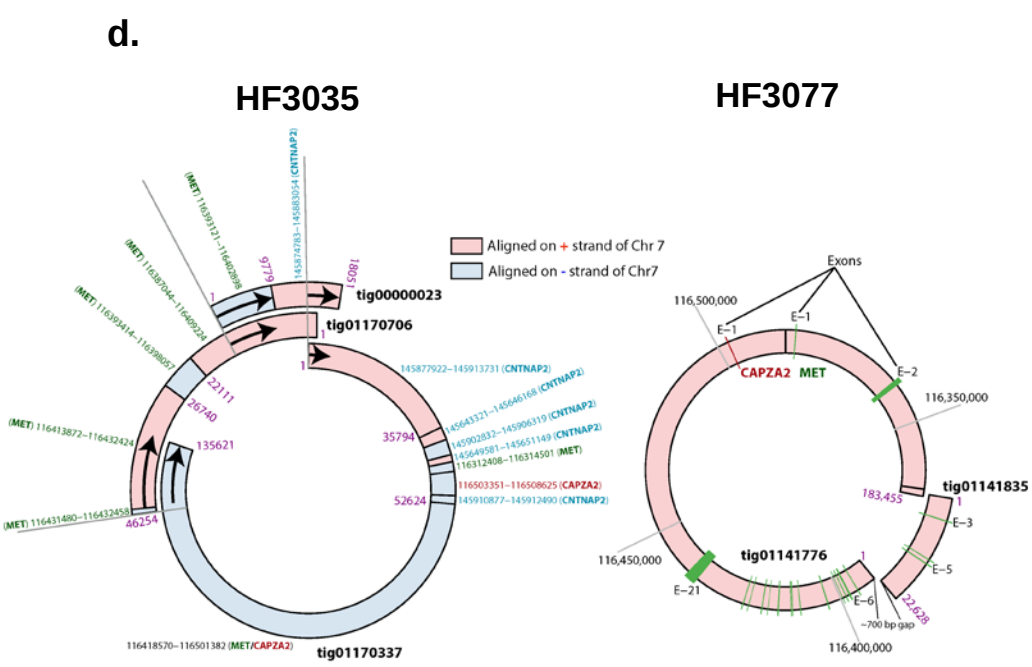


Figure 3. *CAPZA2-MET* fusion. a. The 7q31 locus in three sets of GBM tumors and derivate models. **b.** Coverage-controlled sSNVs detected using exome and deep sequencing (top panel). Color reflects cellular frequency estimates. Bottom panel shows clonal tracing from HF3035 and HF3077 parent tumor to neurospheres and PDXs. Each line represents a group of mutations computationally inferred to reflect a subclone.





***Note**: All genomic coordinates are on chromosome 7.*

Figure 3. *CAPZA2-MET* fusion (cont'd). c. Representative fluorescent in-situ hybridization images for MET (green) and chromosome 7 centromere probes (7qCtr, red) labeling of HF3035 and HF3077 tumor, neurosphere (NS), and xenografts (PDX), and neurospheres established from HF3035 xenograft tumors (PDX-NS1). Passage numbers are indicated for neurosphere cultures. White arrows point to 2 fragmented MET signals in one chromosome in HF3035 samples (2SM). Yellow arrows point to double minute MET in metaphase nuclei of HF3035 neurospheres. The percentage of nuclei presenting MET amplification for each sample is shown. In HF3035, a PDX derived low passage neurosphere was also included. **d.** Double minute structures predicted with long read sequencing in HF3035 and HF3077 xenografts.

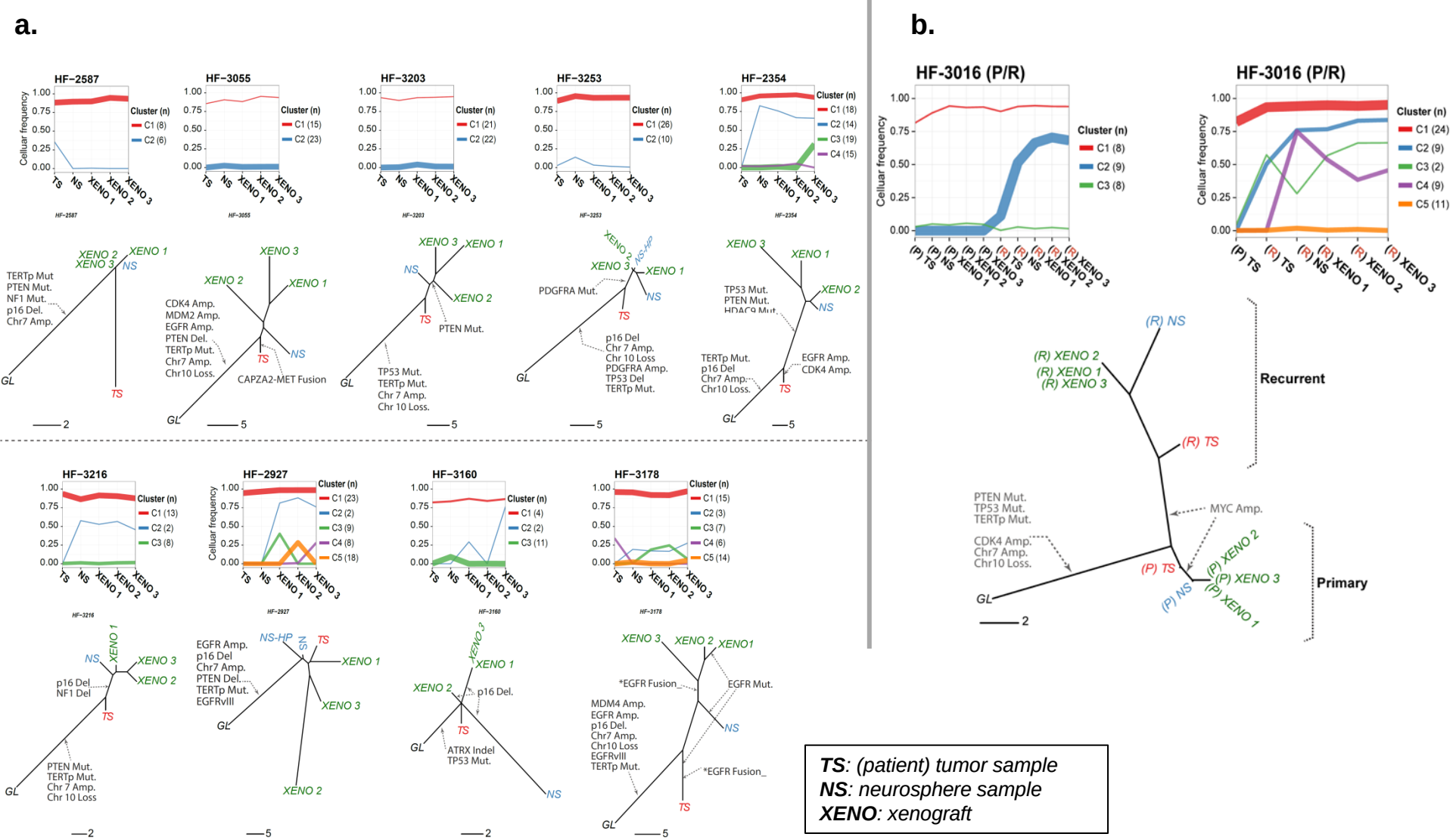


Figure 4. Clone tracing from parent tumor to model systems. a. Clonal tracing of nine primary GBM, their matching neurospheres, and xenografts. For HF3035/HF3077, see Figure 3b. **b.** Clonal tracing of a pair of primary-recurrent GBM, their matching neurospheres, and xenografts.

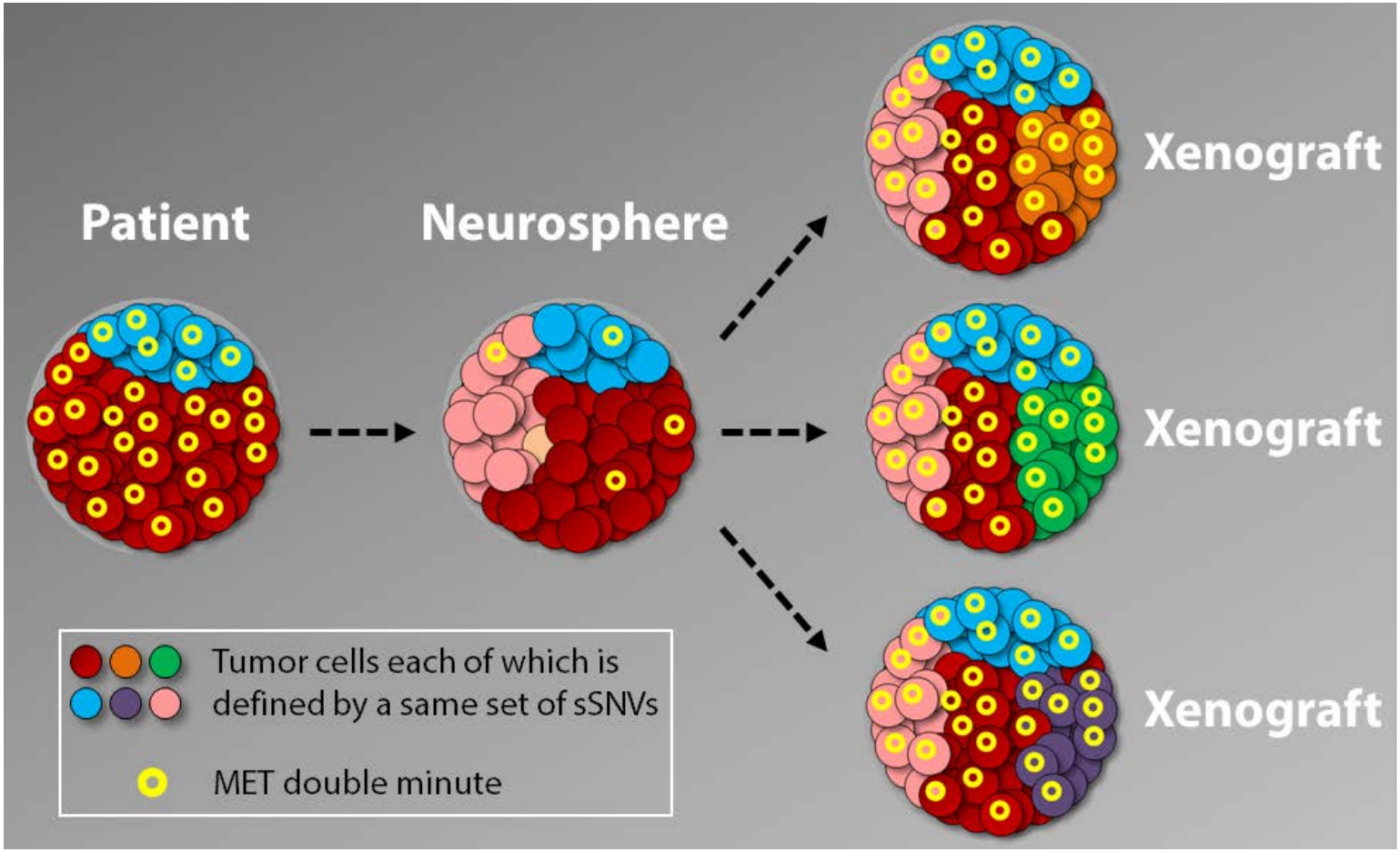


Figure 5. Schematic illustration of double minute driven clonal evolution. The proliferation pattern in samples in which double minutes provide the dominant evolutionary force.