UNIVERSITÀ DEGLI STUDI DI GENOVA

PhD Course in Biotechnologies in Regenerative Medicine

XXXI Cycle

Academic year 2017-2018

Clonal tracking of Glioma progression

Supervisor:

Paolo Malatesta, PhD,Prof

PhD Candidate:

Davide Ceresa, MSc

Contents

Introduction

Gliomas

Gliomas are a group of brain tumors composed by cells that display a phenotype similar to glial cells. The heterogeneity of the different glial phenotypes is reflected in that of gliomas, which include tumors primarily composed of astrocytes, oligodendrocytes, choroid plexus cells or radial glia cells [1].

Gliomas represent almost 80% of all primary brain tumors, they are still mostly incurable and related to high mortality: the most common glioma and the most aggressive form, Glioblastoma Multiforme (GBM), is fatal within two years from diagnosis in more than 70% of cases [2].

The low survival of patients affected by these tumors is a direct consequence of their location, mainly due to the extreme delicacy of the organ from which they originate (in fact, the brain undergoes significant damage simply from the pressure produced by the neoplastic growth), but also due to the scarce accessibility to tumor, which is hardly operable and unreachable by hydrophilic molecules/macromolecules through to the blood-brain barrier (BBB), hindering the use of many of the most biologically active drugs [3].

One peculiarity of gliomas is that, despite the high invasiveness of their most malignant forms, they are usually not prone to metastasize, and only a small percentage (between 0.2% and 0.5%) give rise to metastasis outside the brain [4]. This fact can be partially explained by the low survival

of patients, which does not allow the acquisition of long-range migrating abilities by a mere deprivation of time. However, other causes contribute to this feature, such as the tightly regulated interchange between brain and blood stream by BBB, which avoids extravasation of cells, and the recently discovered lymphatic system of the brain, mainly managed by glial cells [5]. Another important reason is linked to the composition of the brain extracellular matrix (ECM), which lacks of the major glycoproteins, such as collagen and fibronectin, and for this reason is more permissive than other organs [6]. In fact, glioma cells are able to migrate in their native parenchyma, but given the absence of selective pressure, they rarely acquire the ability to engraft and invade organs shielded by connective matrix [7].

Since glioma progression rarely implies metastatization, which is an highly selective process [8], and is located in a relatively permissive environment, it is reasonable to hypothesize that glioma cells are subjected to a minor selection and therefore glioma progression can be easier for a transformed cell than the progression of other tumors. This may reflect and explain glioma intra-tumoral heterogeneity.

Glioma Heterogeneity

In general, the heterogeneity of tumors consist of inter-tumoral and intra-tumoral heterogeneity [9]. Inter-tumoral heterogeneity refers to the differences between tumors from distinct patients, while intra-tumoral-heterogeneity refers to the phenotypical and genomic variations found in different cells within the same tumor. Both inter-tumor and intra-tumor heterogeneity are hallmarks of gliomas, and in particular of glioblastomas [10].

Inter-tumoral heterogeneity

The inter-tumoral heterogeneity of gliomas has been historically tackled by searching an efficient classification for these tumors. This interest is primarily clinical and arises from the necessity of predicting the outcome of gliomas, which may be very different [2], and therefore guide clinical choices due to different risk-benefit ratios.

The first type of glioma classification is based on histo-pathological analysis, and evaluates the tumor histology as well as the differentiation level of the tumor composing cells. This classification is regulated by the World Health Organization (WHO), which defined glioma subclasses and, over time, have been refined thanks to new knowledge on the subject. In the first edition of the classification, published in 1979, gliomas were classified by using parameters concerning the sole appearance of the tissue: according to this classification, gliomas are divided into astrocytomas, oligodendrogliomas or oligoastrocytomas, alluding to the type of glial cell they most resemble (Figure 1) [11].

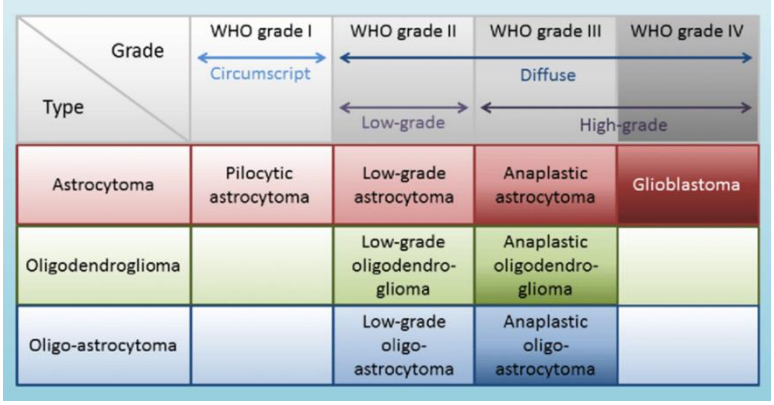

Figure 1: main histological classes of gliomas.

This initial approach was mainly due to the analytical possibilities at the time, since microscopy was the most powerful technique to retrieve qualitative information from tumor biopsies [12, 13]. However, this approach is based on the rationale that different tumors originate from distinct cell types [14]. This view is mostly accepted but in the past decades it has been refined and it is still debated, because even if cell lineage has a strong impact on tumor physiology [15], some lineages are more distinctive than others. In hematopoietic malignancies, for example, tumor biology is strongly related to their cell of origin [16, 17], but in other cases, such as gliomas, the recognition of cell of origin is far from univocal. The first hypothesis on the origin of gliomas was that they originate from mature glial cells that during tumor progression undergo a process of de-differentiation [18], but more recent stronger evidences suggest that staminal or progenitor populations are the most likely cells of origin of gliomas [19-22]. However, the complexity of neural stem cell lineages in the brain [23] makes the recognition of glioma cell lineage even more elusive. Gliomas developed in distinct brain regions can share molecular signatures even if histologically indistinguishable [24], and In the case of ependymomas, distinct tumor signatures resemble those of neural progenitor cells in the corresponding regions of central nervous system [25]. Taken together, these evidences suggest that glioma heterogeneity can be due to their derivation from different site-specific cell populations. Other evidences on the impact of cell of origin in the glioma biology come from mouse models: molecular lesions can be driven in restricted cell populations of recombinant mice by the use of lineage-specific promoters to drive the expression of oncogenes, Cre recombinase [26], or TVA receptor to utilize the RCAS-TVA system [27]. These studies defined that immature populations of brain, but not mature glial cells, are able to give rise to tumors after first hits of gliomagenesis such as *Ras* and

Akt activation [20], or *p53*, *Nf1*, *Pten* [21] and *Ink4a-Arf* [22] inactivation alone. Animal models also highlighted some discrepancies: Although vast majority of studies identifies radial-glia like neural stem cells (NSCs) as progenitors for most gliomas, some studies identified other populations, such as reactive astrocytes for PDGF and NF1 driven glioblastomas [28], and several reports suggested that gliomas can arise from oligodendrocyte precursor cells (OPCs) [29-31]. In particular, two works studied glioblastomas driven by the same molecular lesions in different cells, and concluded that both suppression of *p53*, *Nf1* and *Pten [32]* , and overexpression of *PDGF-B [33]* in OPCs are able to induce glioblastomas, but they belong to different subclasses if compared to glioblastomas derived from $Nestin^{+}$ NSCs. Strikingly, under extreme condition such as overexpression of Polyoma virus middle T antigen, even mature astrocytes can give rise to mixed astrocytic and oligodendrocytic gliomas [34]. Interestingly, this last study showed that glioma cell lineages are plastic enough to be generated by a single cell type. Consistent with this evidence, other researchers demonstrated that OPCs be cell of origin of both astrocytoma and oligodendroglioma in an oncogene-dependent manner: overexpression of PDGF-B leads to oligodendroglial tumors, while combined overexpression of KRAS and AKT generates gliomas displaying astrocytic histology [35]. This result shows that molecular lesions can be dominant to cell of origin in the determination of glioma phenotype.

Even if lots of evidence show that glioma inter-tumoral heterogeneity is a consequence of the cell of origin biology, its recognition is far from being easy, and other factors such as mutations can have an equal or even stronger impact on their biology [19, 36]. An alternative approach to obtain insights on the inter-tumoral heterogeneity of gliomas and is to classify them by their mutations and/or their gene expression: In particular, gene expression profile can give integrated

information about cell of origin and genomic lesions. Using this paradigm, Glioblastomas have been clustered in base of molecular lesions, focusing on the repercussion of unregulated genes on known signaling pathways. In this way, it is possible to identify three subgroups characterized by *(1)* an increased signaling of EGFR, *(2)* an increased signaling of PDGFRα, or *(3)* a decreased signaling of NF1 [37].

A first transcriptomic clustering of glioblastomas was performed by k-means clustering of 76 GBM samples, and identified 3 subgroups, named them Proneural, Proliferative, and Mesenchymal by recognizing the dominant features of the gene list that characterize each subclass. Proneural subclass display a better prognosis than the other two, and interestingly, most recurrences shift to the mesenchymal subtype, the most aggressive. This classification proved its prognostic value correctly predicting outcome in patients depending on their belonging to the three subclasses [38]. Few years later, a similar analysis defined one of the most important subclassifications of glioblastomas, comparing the transcriptome of 433 Glioblastomas obtained from *The Cancer Genome Atlas* (TCGA). Differential gene expression analysis revealed 840 modulated genes able to divide this single WHO class in four subclasses, termed classical, neural, proneural and mesenchymal [39]. Consistent with the other clustering, neural and proneural tumors have a better prognosis than classical and mesenchymal glioblastomas and furthermore, these subtypes display different responses to treatment: proneural tumors are less aggressive but also less responsive to strong therapies compared to classical and mesenchymal tumors. Therefore, this classification can accurately guide the therapy of patients.

The importance of genomic and transcriptomic classifications, as well as the interest on this approach, started to grew since this classification and, in 2016, WHO classified tumors of the central nervous system considering both histopathological characteristics and genomic characteristics. The categories are then remodeled and more objective then the previous classifications.

Intra-tumoral heterogeneity

Cells which compose the same tumor can be genetically and/or phenotypically different each other [40, 41]. This phenomenon is called intra-tumoral heterogeneity, and it has shown to be the main cause of the partial response to therapies and the development of recurrences [42]. For all these reasons, intra-tumoral heterogeneity received a lot of interest in cancer research.

Gliomas are tumors characterized by a high intra-tumoral heterogeneity [43-45]. Multi-region sampling of human glioblastomas revealed that different regions of same tumor can belong to different molecular classes [46], and more strikingly single-cell RNA sequencing confirmed that most Glioblastomas are composed by cancer cells that individually belong to all the four molecular GBM subclasses. These data also demonstrate that cancer cells composing same glioma can be different each other as much as different gliomas.

While in principle inter-tumoral heterogeneity could be explained by the existence of different cells of origin, intra-tumoral differences are deeply correlated to tumor progression and evolution. Since 1976, it has been proposed that tumor evolution imply multiple branching steps leading to the emergence of subclones which can expand and sometimes eradicate or sometimes coexist with other clones, generating a dynamic and heterogeneous population of tumor cells

[47]. This model is ongoing widely accepted, and genomic analyses of heterogenic tumors reveal that in many cases it is possible to reconstruct the evolutive process of tumor progression, and find the single branching points leading to the intra-tumoral heterogeneity [48-50] . This approach also lead to the evidence that metastasis, usually phenotypically different from the bulk primary tumor site, derive from a clone which already diverged and expanded in the primary tumor site [49].

Intra-tumoral heterogeneity can be addressed by clonal tracing the tumor cells by the use of genetic barcoding.

Genetic barcoding

Genetic barcodes (or DNA barcodes) are short sequences of DNA in the genome of a cell that can univocally tag the cell itself and consequentially its whole progeny. Genetic barcodes can be either natural or synthetic sequences.

Natural DNA barcodes are usually short stretches of coding sequence that are used to identify different species in heterogeneous populations. The discovery and use of natural barcodes is not trivial: to be useful, a genetic barcode needs to be variable enough through evolution to make species recognizable, but in the same time stable enough to be retrievable in a big group of organisms. Examples of successful barcodes are *Cytocrome C Oxidase I* for animals [51], and chloroplast DNA for various plant species [52], but the most successful and used natural DNA barcode is a region of ribosomal 16S sequence, shared by almost all prokaryotes, that can identify a very large set of bacteria and archaea species [53], making it extremely useful for ecological [54], and medical microbiome studies [55].

On the other hand, synthetic DNA barcodes are synthetic DNA sequences that can be designed to be as heterogenic as possible, but need to be integrated in the genome of the cells in order to label them. First implementations of this approach used microarray-based recognition of barcodes, and consequently barcodes were designed as 100 bp long sequences to allow efficient hybridization [56]. This method made possible to study clonal dynamics of T cell lineage [56, 57]. However, it was extremely labor-intensive and implied the use of mixtures of pre-defined DNA sequences, making it difficult to scale up the number of different barcodes. For this reason, DNA barcoding did not spread much before the establishment of Next Generation Sequencing (NGS) techniques, which allow the parallel sequencing of billions short sequences [58]. Thanks to NGS, barcodes can be identified and quantified without any prior information, making the use of high complexity barcode libraries more practical. Since then, synthetic barcodes become shorter, usually not much longer than 25 nucleotides, given that a completely degenerated sequence of this length can have more than 10^{15} possible forms and thus can be useful for the study of extremely heterogeneous populations [59]. Genetic barcodes have been extensively used to track the clonal dynamics of hematopoietic system [60, 61], and cancer [62-65].

Synthetic barcodes design is highly permissive, and indeed lots of variations have been developed. For example, modular barcodes containing multiple CRISPR/Cas9 target sites have been designed to progressively accumulate mutations during time enabling the tracing of divergent populations in vivo [66, 67], while Brainbow mouse-inspired [68] approaches led to the design of single modular barcoding loci that can be integrated in-vitro or in-vivo and generate more than 10^4 [69] or 10^9 [70] different sequences after *Cre-Lox* recombination, and can be extremely useful for *in-vivo* studies.

Rationale and Aim of the thesis

The clonal dynamics of glioma progression is not fully understood. Most of the information about tumor progression is derived by post-analysis of already progressed tumors and reconstruction of clonal evolution from already progressed tumors.

The model of gliomagenesis developed in our lab is induced by the overexpression of PDGF-B, a known oncogene typical of glioblastomas, in embryonic progenitor cells by in vivo injection of retroviral particles in telencephalic ventricles of mice embryos at E14 stage. Transduced cells grow and generate neoplasms in mice brains that undergo tumor progression and, after a first low-grade stage, they gain malignancy and resemble human Glioblastoma [71]. Since in this model gliomagenesis is induced by retroviral-mediated gene transfer, is particularly suitable for clonal tracing using genetic barcodes. The oncogenic retrovirus can be modified inserting a degenerated barcode sequence generating a viral library that can, in a single step, mimic the same first hit of gliomagenesis in multiple cells and label each transformed cell, allowing the clonal tracing of each cell composing gliomas.

The aim of this thesis is to highjack this mouse model to directly analyze the clonal dynamics of glioma progression and investigate eventual clonal bottlenecks and relation between clones at different malignancy stages.

Results

Barcoding library production and validation

General design and plasmid library construction

I decided to insert the barcode (BC) sequences in the expressed retroviral region containing PDGF-B and GFP (Figure 2A). BCs will be then integrated in the cell genome , but also transcribed, thus giving the opportunity to recover information either trough genomic DNA extraction or RNA extraction and retro-transcription. Given that the BC insertion in the vector must be as efficient as possible, I initially modified the vector by inserting, via NotI cloning, a landing cassette containing a stuffer DNA sequence flanked by two short regions. After the final step of library production these regions will be flanking the BC and will be used to amplify them by an efficient and specific PCR.

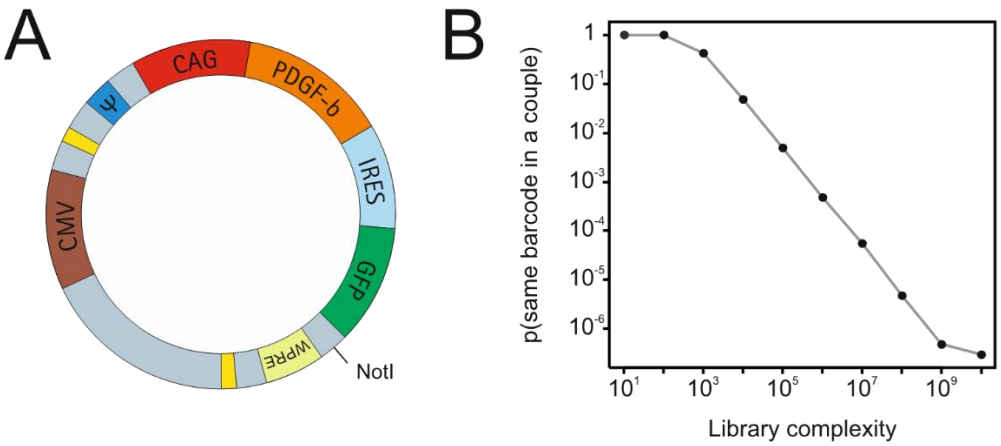

Figure 2: Barcoding vector design. A: schematic view of the tumorigenic vector and barcode insertion region, highlighted by the NotI site. B: probability to label two cells with the same barcode derived from silico simulation of infections using differently complex libraries.

To define the flanking regions sequences I generated three pairs of synthetic primers, complementary to a random *in-silico* generated template. Primers were tested on murine genomic DNA and cDNA derived from murine NPCs and PDGF-B driven tumors. One pair resulted to be completely orthogonal to the biological samples and was chosen for the vector design (Figure 3).

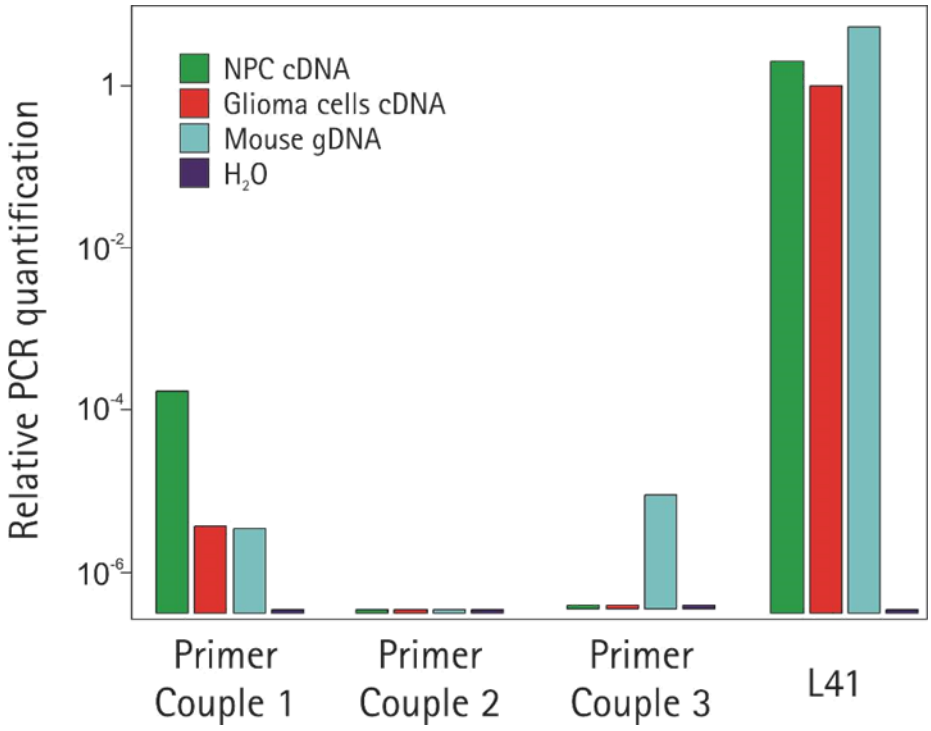

Figure 3: Synthetic primers specificity test. Although their specificity had already been tested bioinformatically by BLAST algorithm, they were used in qPCR on murine samples in order to validate them empirically. Primers for the housekeeping gene l41 were used as a positive control

Barcode sequences were obtained by organic synthesis of 22-nt long degenerated oligonucleotides, consisting of 11 "B" (T, C or G) and "H" (A, T or C) repeats. This degenerated pattern can be used as quality control and has $3^{22} \sim 3x10^{10}$ possible forms, allowing a huge theoretical complexity.

In order to minimize the probability to tag different cells with the same barcode, it is necessary to generate a sufficiently complex barcoding library, harboring enough different barcode sequences. In-silico simulation revealed that to keep this probability under 5% or 1%, the library must harbor respectively around 10^4 or 10^5 different sequences (Figure 2B), which is so my target complexity.

Barcode sequences were inserted in the viral vector by Gibson Assembly and 150 ng of the reaction product (about 10^{10} molecules) was electroporated in high-efficiency *E coli* cells. In order to avoid statistical drifts and inequalities in library composition, bacteria were grown in high volume (5 l) of LB and their growth was stopped in exponential phase. I retrieved 700 µg of plasmid DNA library, that was initially sequenced by Sanger sequencing to have a first estimate of the heterogeneity of the barcode region: the electropherogram obtained confirms the presence of a degenerated 22-nt long stretch (Figure 4).

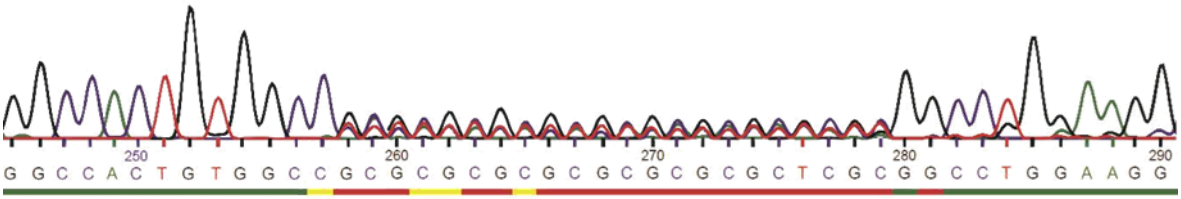

Figure 4:Sanger sequencing of the barcode region. Consistently with the imposed barcode pattern, all the 22 positions of the barcode contain three peaks, revealing a degenerated sequence.

Plasmid library validation

NGS sequencing of the plasmid preparation was performed to better evaluate the heterogeneity of the library and to set up the BC sequencing protocol. I chose to start using Ion torrent platform because it allows relatively low depth (about $5 \cdot 10^5$ reads) sequencing, perfect for the validation of a single sample. Two independent sequencing libraries were generated by PCR and BCs were retrieved by bioinformatic alignment of the viral backbone. I analyzed the nucleotide composition and I verified the BH pattern (Figure 5).

Figure 5: Sequence logo generated from all barcode sequences retrieved from NGS sequencing. The pattern consisting of BH repetitions is recognizable

All the different BCs were clustered basing on Hamming distance: as expected from in-silico simulations, barcodes have a mean Hamming distance of 14 nucleotides, and a clearly different population of smaller Hamming distances is recognizable, mostly due to PCR and NGS errors: this first observation helped me to set the threshold to distinguish genuinely different barcodes from PCR/NGS artifacts.

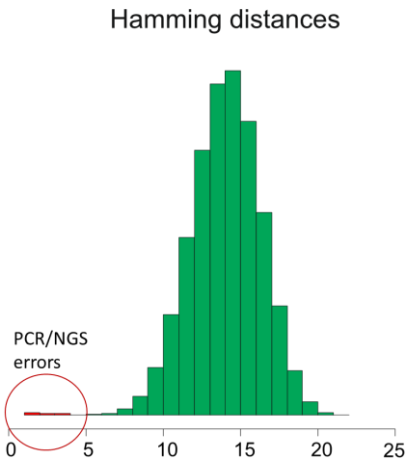

Figure 6: distribution of Hamming distances between barcode sequences. A distinct subpopulation of low distances is recognizable and can be explained by PCR mutations or NGS errors

Once clustered, BCs having Hamming distance >5 were counted in both subsampling of the plasmid library. $2*10^4$ different sequences have been found in each sequencing, and strikingly

only 138 sequences are shared by the two, leading to an estimate of more than 10^6 different sequences in the total library.

Retroviral library construction and in vitro validation

Once the plasmid library was validated, high titer (10^6 infecting particles/ml) library of retroviral vectors was generated by transfecting the plasmid library into packaging cells. Before using this library for in vivo experiments, I had to perform preliminary in vitro tests to check that:

1. The viral library is highly complex and each barcode sequence is equally represented, and so able to label each infected cell with a different barcode (barcoding efficiency).
2. Each infected cell integrates just one of the two retroviral chromosomes and contains just one barcode sequence (unique labeling).

Barcoding efficiency

I performed 5 independent infections of NIH-3T3 cells using the viral library at low multiplicity of infection (MOI). 12h after the infection cells were detached and divided in two samples: 10% of them was plated at low density in glass coverslip to estimate, after some cell divisions, the number of infected cells by epifluorescence microscopy thanks to the GFP expression. The remaining 90% of them was plated and collected after 3 cell divisions, the genomic DNA was extracted and amplified by PCR to generate 5 Ion Torrent sequencing libraries.

The number of GPF positive clones estimated by fluorescence microscopy nicely corresponds to the number of different barcode sequences retrieved from Ion torrent sequencing, and no particular imbalance in barcode composition was found.

I used the same data to estimate the viral library complexity. I counted, for each sequence, the number of different infections in which it was detected, and I compared the experimental data to different simulations based on different defined complexities: by using as parameters the frequencies of barcodes found in 1, 2, or 3 independent infections I have a solid estimate of the order of magnitude as 10^4, and in particular I estimate at least $1.5*10^4$ different sequences.

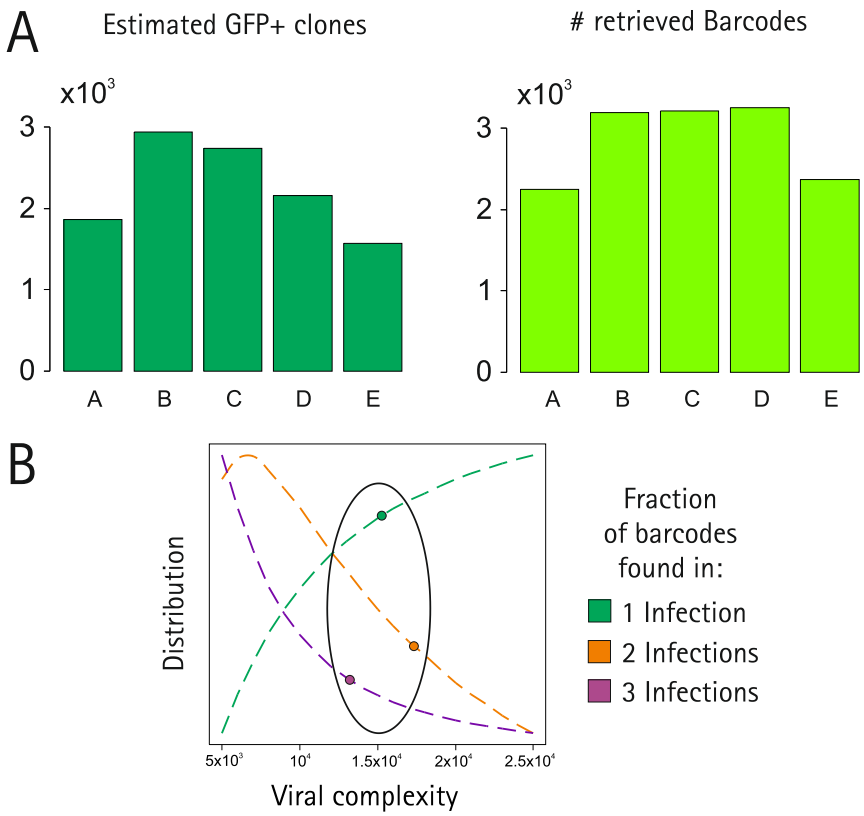

Figure 7: barcoding efficiency validation and heterogeneity estimation. A: number of estimated infected cells by fluorescence microscopy compared to the number of detected barcode sequences. B: experimental data about barcode distribution between the 5 independent infections is compared to viral complexity simulations and consistently indicates the presence of 1.5+10^5 different sequences.

Unique labeling

To observe the multiplicity of barcode integration in uniquely infected cells I infected NIH-3T3 cells as in the previously described experiment and obtained clonal colonies of GFP positive cells by limit dilution cloning. I extracted the DNA from 10 clones and analyzed the barcode region by Sanger sequencing: each clone harbored a single barcode sequence (Figure 8:), suggesting that usually just one viral chromosome is integrated in each cell.

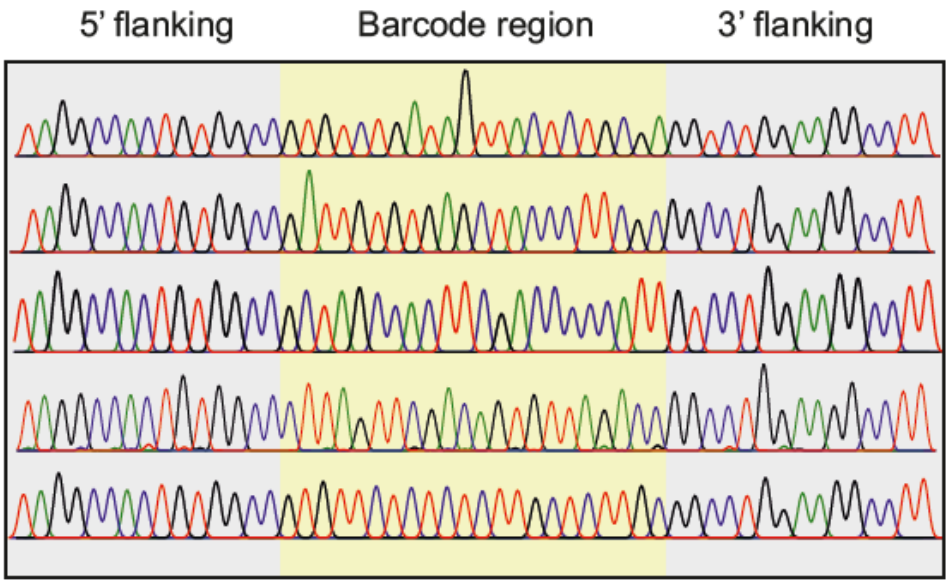

Figure 8: sanger sequencing of barcodes from genomic DNA of clonal progeny of infected cells. Each clone harbors a single barcode sequence.

Tumoral progression in primary tumors

Embryonic infections

I then proceeded with the in vivo experiment: the retroviral library was injected into the telencephalic ventricles of E14 murine embryos. The infection is not lethal for the mice, which complete che development and are born. It has already been shown that at this stage the PDGF-B overexpression is sufficient to induce neoplasms in the brain that undergo tumor progression and pass through a low malignancy stage, in which they are composed by cells that are not tumorigenic if orthotopically transplanted in syngeneic mice (low grade gliomas), and by time reach a more malignant stage gaining the ability to give rise to secondary tumors (high grade gliomas).

I collected the brains from 7 mice embryos at stage E18, 4 days after the viral infection, to analyze the starting heterogeneity of the transduced cells. Given the low amount of GFP positive cells I decided to extract the genomic DNA from all the lysed brain. The barcode region was

amplified and sequenced by pooled Illumina sequencing as described in materials and methods. For each embryo, I generated 3 independent sequencing libraries, in order to normalize the possible inequalities due to PCR amplification and to estimate the sample complexity by studying the amount of barcodes shared by different subsampling. The initial amount of in vitro infected cells is $20 \pm 5 \cdot 10^3$ (Figure 9).

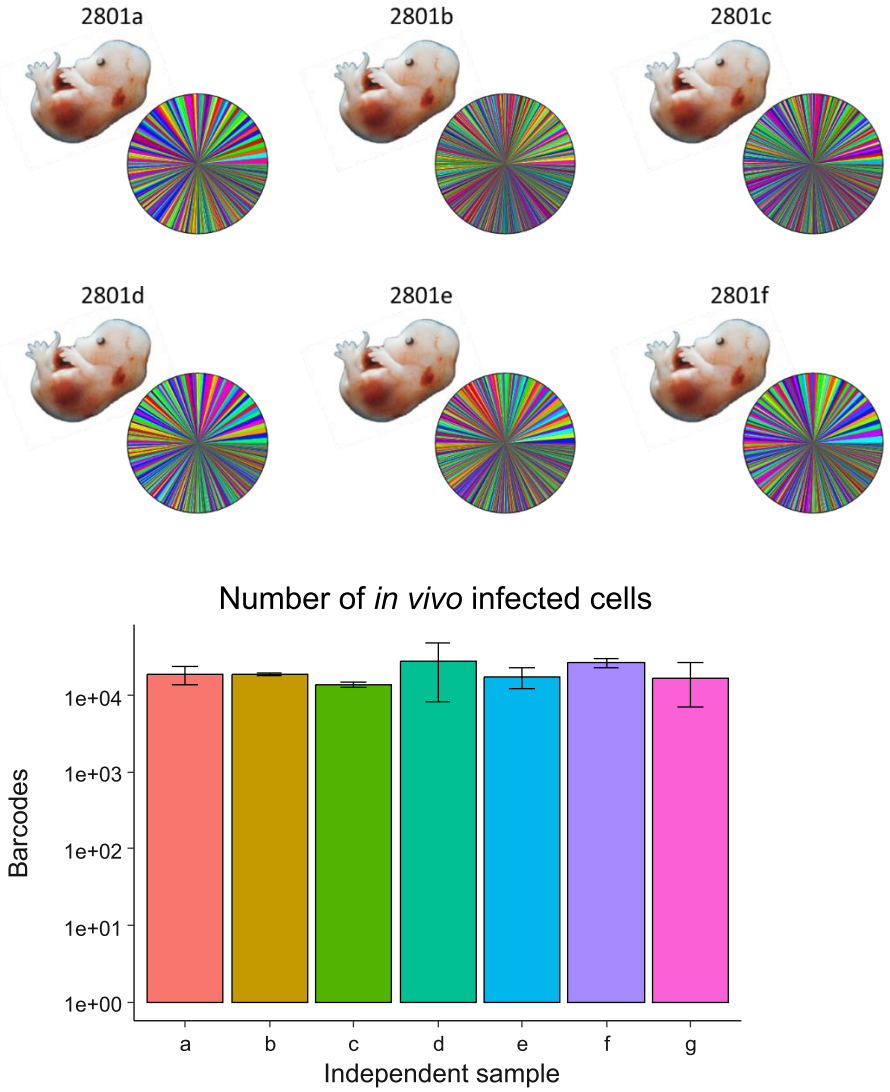

Figure 9: Barcode sequencing of embryonic brains. Each embryo contains evenly distributed barcodes that can be estimated at different precision for each sample. All the infections lead to the same order of magnitude of transduced cells.

Progression from low to high grade gliomas

I then dissected tumor tissue from the brain of mice harboring low-grade tumors. GFP positive population were enriched by FACS sorting and DNA was extracted from sorted cells. I decided to generate 3 sequencing libraries from each low grade glioma, as I did for the injected embryos.

NGS sequencing reveals that even in earlier stages of glioma progression the initial population goes through a strong bottleneck: tumors harbor hundreds or even tens of different barcodes, suggesting that just a small percentage of transduced cells had the ability to reach the low-grade tumoral stage. Interestingly, the distribution of clones in the total tumoral mass appears to be quite homogeneous in most tumors, but some clones are bigger and in all the cases about half of the total mass is composed at most by the ten bigger clones, while the other half tumoral mass is usually composed by tens or hundreds smaller clones (Figure 10).

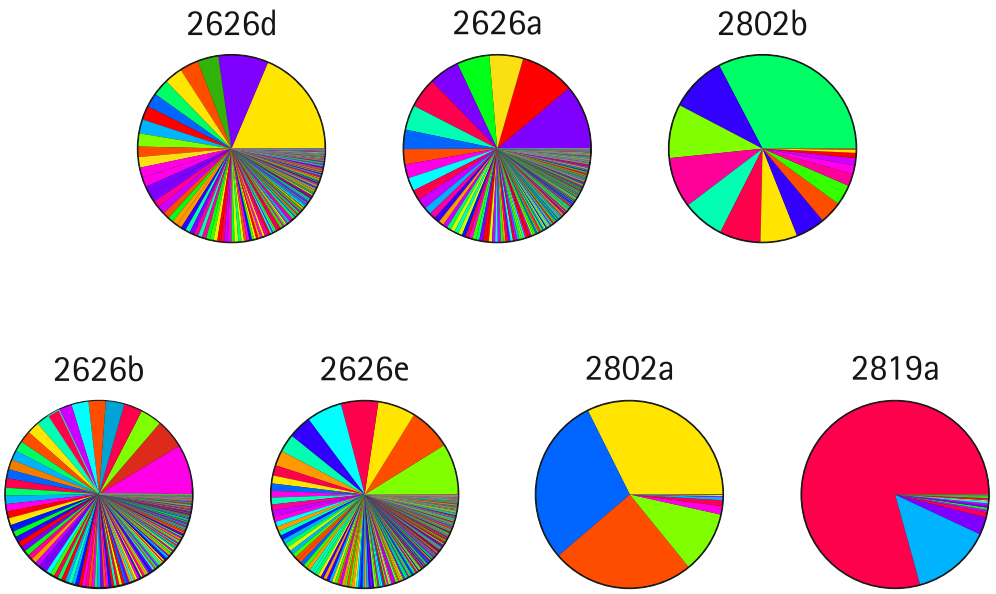

Figure 10 :clonal composition of low-grade tumors. Clones are represented from bigger to smaller.

High-grade tumors, on the other hand, are characterized by a much lower clonal heterogeneity: each tumor contains on average 20 total clones, and more strikingly one or two main clones represent the main percentage of tumor mass. By this clonal analysis, I can also divide the High grade stage into 2 categories: tumors retrieved until 120 from embryonic

transduction are usually oligoclonal, while later tumors are mainly bi-clonal or even monoclonal (Figure 11).

These data suggest that the tumoral progression, from the first molecular lesion to the last high-grade stage, is very unlikely. Given the number of initially transduced cell and the number of clones composing High-grade gliomas. This probability can be estimated as 10^{-4}. A model with such probability imply that in a consistent proportion of cases not even a single clone manages to gain full malignancy. However, The penetrance of PDGF-B transduction is 100%. This evidence suggest that this estimate is the lower limit of the real probability of glioma progression, and other phenomena, like internal competition of clones, can explain the low amount of clones in full-blown gliomas.

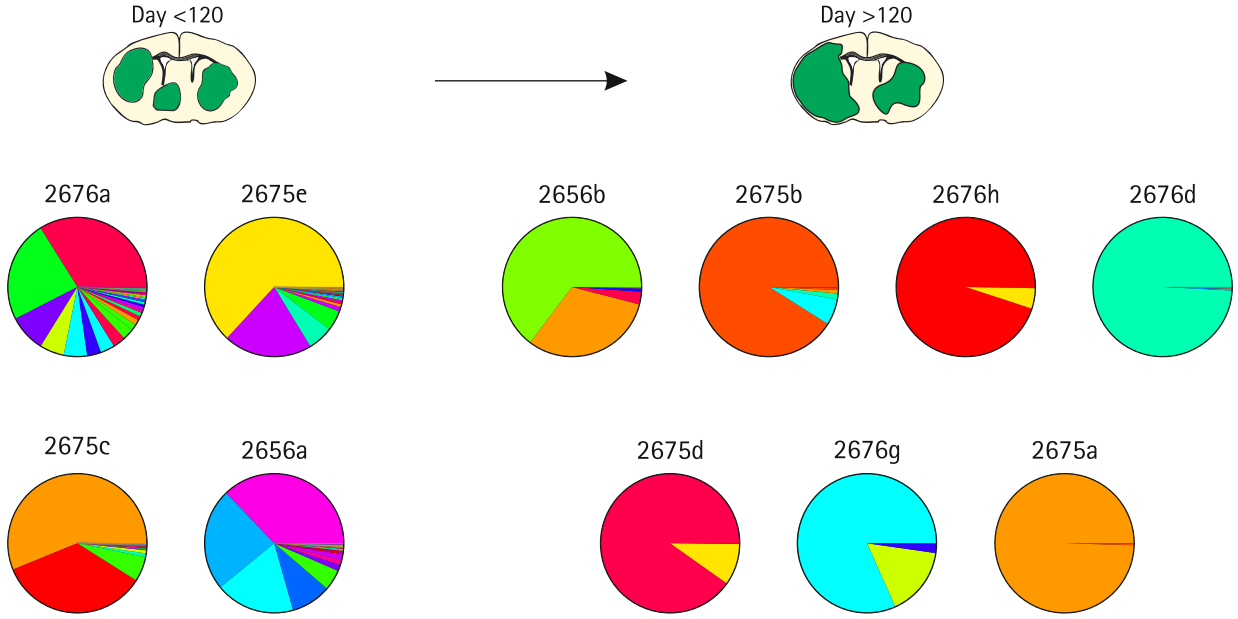

Figure 11: clonal representation of high-grade gliomas, divided by the time of onset. Clones are represented from bigger to smaller.

To better understand the clonal dynamics of glioma progression, I studied the relation between clonal heterogeneity of gliomas and their time of onset, which can be considered the time of tumoral progression. The clonal complexity of gliomas dramatically decrease over time: If the probability of clone loss is equal in the tumoral progression, the expected rate is an inverse

exponential. When plotted in a logarithmic scale, however, the number of clones shows a statistically significant, non linear correlation more than exponential decrease over time (Pearson correlation = -0.79; $p<10^{-5}$, $R^2 = 0.63$) , suggesting that during earlier stages of glioma progression clones are more prone to disappear and become more stable in later stages. This may be the result of the rising number of cells in each clone, but may suggest that glioma cells gain of malignancy make them stronger and less prone to be eradicated by the host. Interestingly, even if not clearly recognizable, the clonal purification rate in the gliomas is continuous and no main bottlenecks are discernable in a particular timeframe (Figure 12 A).

A finer estimate of clonal heterogeneity can be made by Shannon-Weiner H' index, which integrates both number of clones and their relative distributions. This index shows a strong exponential fit with the time of tumoral progression (Pearson correlation = -0.96; $p<10^{-5}$ $R^2 = 0.92$), suggesting that even if clones are progressively more stable in gliomas, their relative differences keep exacerbating with a constant rate until the most malignant clone prevails.

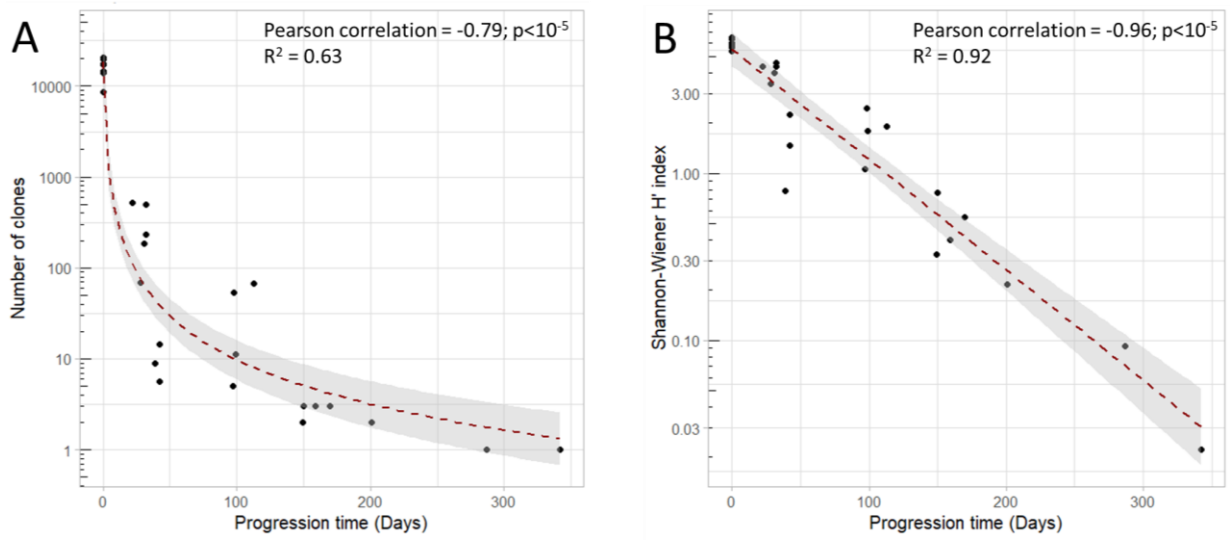

Figure 12: Clonal heterogeneity of gliomas during tumoral progression. Each dot represents a tumor or, in case of time 0, infected embryonic brains. A) The estimated number of clones in each tumor. B) The global heterogeneity estimated by Shannon-Weiner H' index.

Acutely dissociated cells derived from glioma masses are able to give rise to secondary tumors in a onset time-dependent manner. Tumors progressed more than 70 days resemble high-grade

gliomas but they are not invariantly tumorigenic if orthotopically transplanted in syngeneic mice. 0 out of 2 oligoclonal tumors progressed between 70 and 120 days gave rise to secondary tumors (Figure 13 A), while 4 out of 5 clonally homogeneous tumors (progressed more than 120 days) gave rise to clonally consistent secondary tumors. Two of them are monoclonal, and their secondary tumors contain only the parental clone as expected (Figure 13 C). The other two are composed by at least two clones, and in one case the clones are evenly balanced. Strikingly, all secondary tumors are composed by the same single parental clone (Figure 13 B). This is an evidence that high-grade gliomas can contain "passenger" clones, which are not able to engraft in syngeneic mice, and their presence in the primary mass is not sufficient to consider them as tumorigenic.

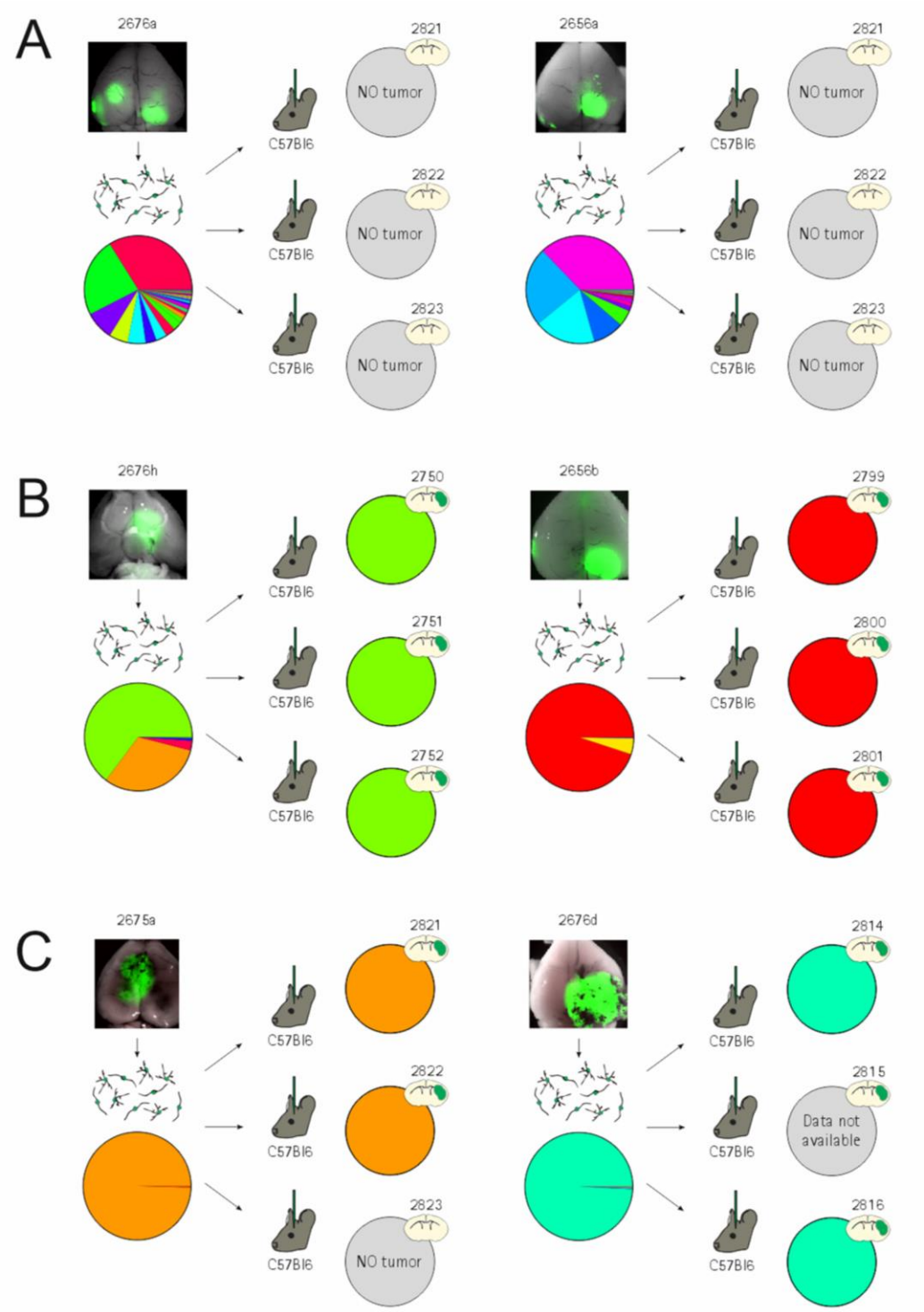

Figure 13. Transplantation assays of cells derived from primary tumors. A: Tumors progressed more than 70 and less than 120 days. B: Oligoclonal tumors progressed more than 120 days. C: monoclonal tumors progressed more than 120 days.

Tumor progression of *in vitro* transduced neural progenitor cells

I then investigated the clonal dynamics also in an alternative induction of gliomagenesis, in which neural progenitor cells are explanted from E14 mice embryo and infected *in vivo* 2 hours after dissociation. After few doublings cells are transplanted in the brain of adult immunocompetent mice and gain malignancy giving rise to tumorigenic gliomas. Such model can be useful to analyze tumoral progression of same pools of early-transformed cells.

4 independent pools of NPCs were infected at low MOI 2 hours after explant and dissociation. Each pool was transplanted in at least two mice and, after glioma development, tumoral masses were retrieved, sequenced and compared to the initial pool of transformed cells. Interestingly, the clonal architecture of these tumors is similar to that of the high-grade embryonic gliomas, suggesting that the two models are subjected to similar clonal dynamics. Some clones are among the most represented in different tumors derived from the same parental NPCs, and in one case (infection 1) 3 out of 4 tumors are composed by the same principal clone (Figure 14). This result suggest that some divergences in the first stages of glioma progression can strongly, but not completely, affect tumor progression.

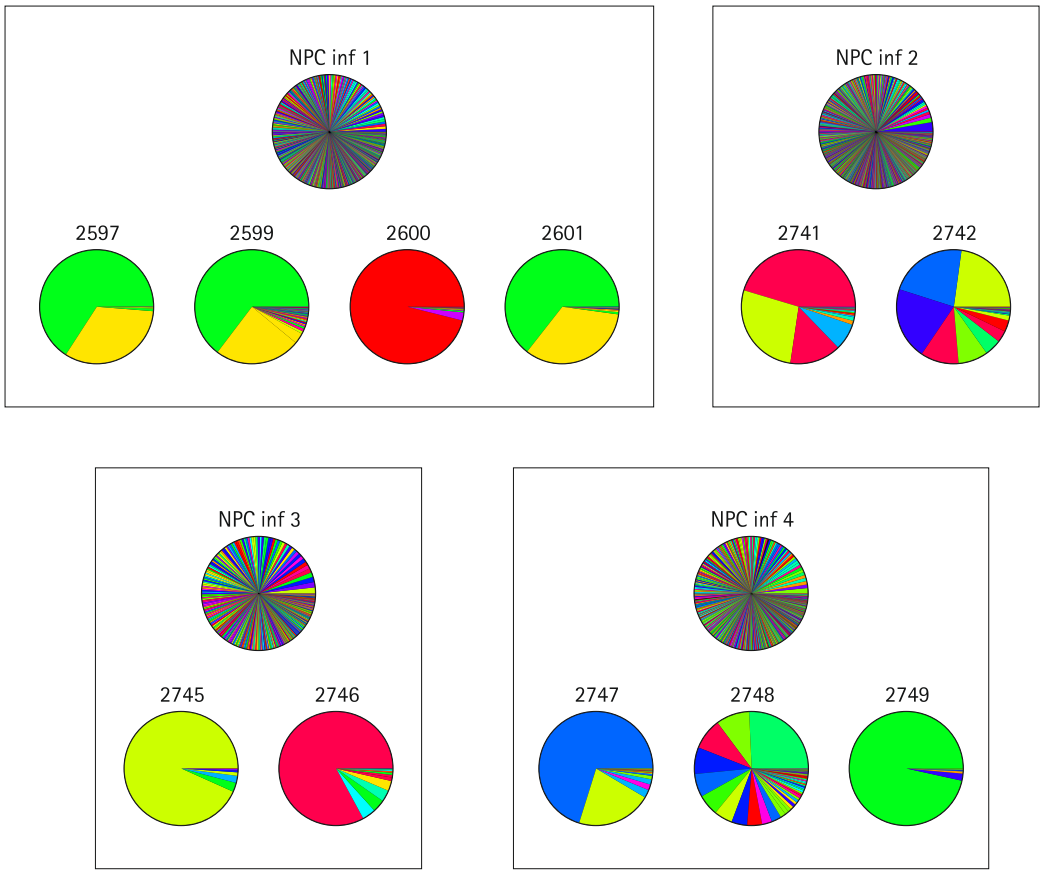

Figure 14: clonal representation of infected NPCs at day of transplantation and derived tumors. Clones are represented from bigger to lower.

Also in this case, the loss of clones is faster than exponential during tumor progression. Tumors start from one order of magnitude less clones than embryonic tumors, but they show an equal, shifted kinetics (Figure 15 A). Their overall heterogeneity, on the other hand, decreases with the same exponential kinetics but faster: tumors become almost monoclonal in less time (Figure 15B).

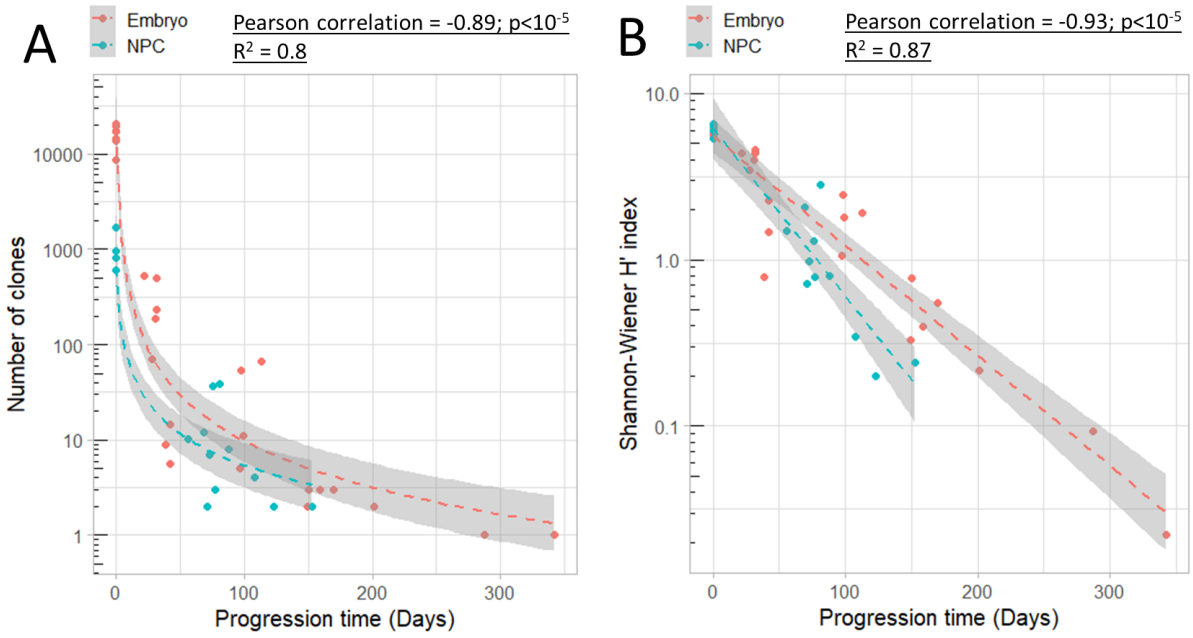

Figure 15: Clonal heterogeneity of gliomas during tumoral progression in the two models. Each dot represents a tumor or, in case of time 0, infected embryonic brains or infected NPCs at day of transplant. A) The estimated number of clones in each tumor. B) The global heterogeneity estimated by Shannon-Weiner H' index.

Glioma progression in immunodeficient mouse

Recently published data from our laboratory [72] demonstrated that although low-grade glioma cells are not able to engraft in immunocompetent animals, they can grow and give rise to secondary tumors when transplanted in immunodeficient mice. Strikingly, cells continue tumor progression even in this model, and in fact, secondary tumors gain the ability to engraft also in immunocompetent mice. I used this knowledge to observe how the same pool of low-grade gliomas undergo tumor progression in independent mice.

GFP-positive cells from a low-grade glioma were transplanted in 4 NOD-SCID mice and 2 C57-bl6 mice as control, and part of them was used to sequence the barcodes of primary tumor. The low-grade glioma cells did not engraft in C57-bl6 mice, but generated secondary tumors in 3 out of 4 NOD-SCID mice. DNA from all the secondary tumors was extracted and sequenced, and cells from 2 secondary tumors were injected in immunocompetent mice. DNA from 3 tertiary tumors was collected and all the samples were sequenced using Illumina platform.

Sequencing data is shown in Figure 16. Each secondary tumor is composed by one major clone that represents more than half of total mass, one or two intermediate clones and some less proliferated clones. By this point of view, progression in immunocompetent mice resembles the progression observed in primary tumors, leading to oligoclonal high-grade tumors. An interesting finding is the agreement among the engrafted clones: although the main clone is different among independent secondary tumors, they derive from just 5 primary clones, which are probably more prone to engraft and keep gaining malignity. 2 of the clones are visible in each secondary tumor, 2 are shared between 2 secondary tumors while the fifth is detectable in just a secondary tumor. A similar concordance, combined with the fact that other clones of equal size have not been able to attract in any animal suggests that already in early stages of tumor progression some clones are partially predetermined to give rise to secondary tumors, and so are more malignant than the others.

This experiment provides also other evidences. Since the same number of cells was transplanted in mice for primary and secondary tumor induction, and clones engrafted in tertiary tumors were a consistent proportion of the initial low-grade tumors, therefore amount of injected cells belonging to particular clones is often similar, although the engrafting efficiency in immunocompetent mice is strikingly different. This finding highlights that tumoral progression is not simply due to preferential expansion of intrinsically more malignant clones, but that clones qualitatively change and become more aggressive. This evidence supports the idea of tumor progression as a gradual gain of malignancy opposed to the idea of a complete early divergence and selective clonal expansion.

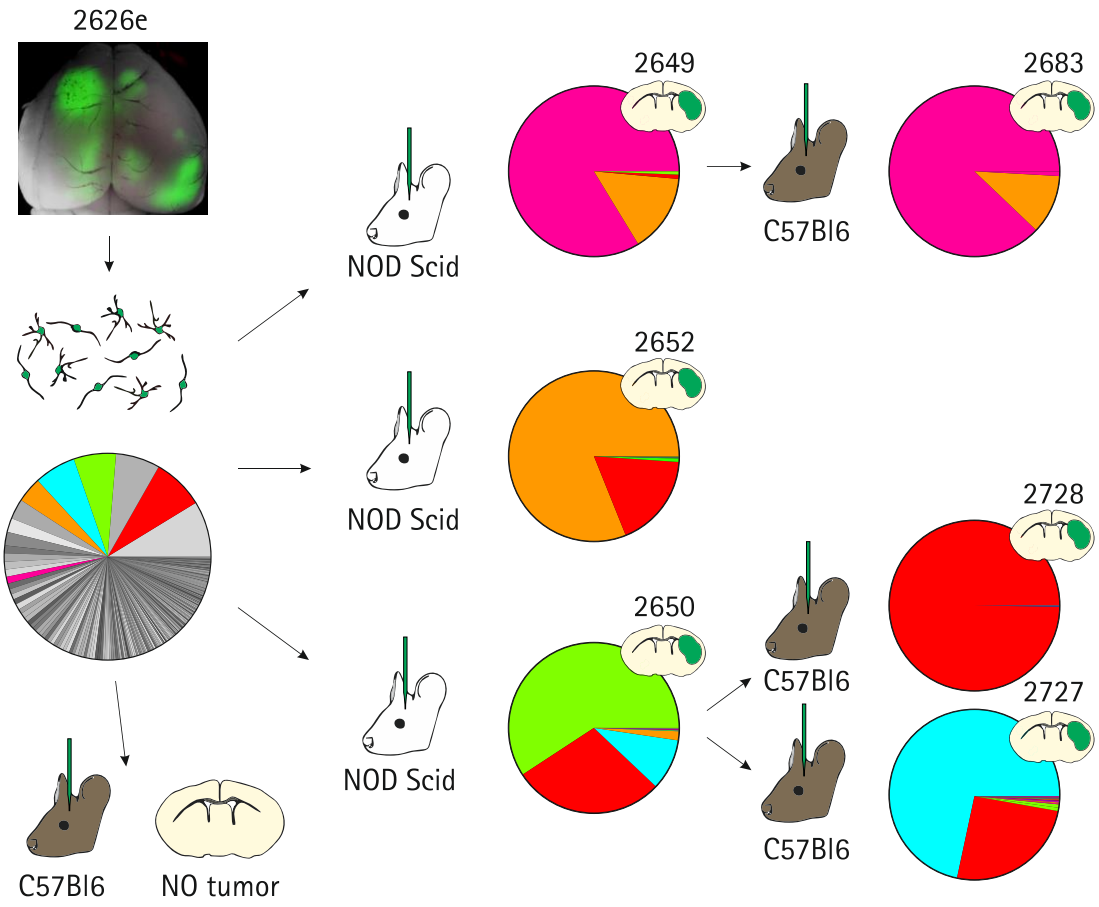

Figure 16: Schematic representation of the experiment and clonal composition of tumors.

Discussion

Studying the clonal heterogeneity of gliomas can help to estimate how many different populations compose them and their relation over time. This information is extremely important in the evaluation of target-specific therapies because can suggest how likely it can be shared between all tumor cells.

The experimental design that I set up allows the clonal analysis of gliomas in different stages of tumor progression. Our data suggest that the entire tumor progression is a very unlikely process and implies strong bottlenecks in the initial pool of thousands transformed cells. This also suggest that even a molecular lesion like the overexpression of PDGF-B, already known to invariantly induce gliomas, is not the main event in gliomagenesis and other critical mutations are still needed for the full malignancy. During tumor progression clones undergo a dramatic selection, and interestingly their probability of tumoral progression is not fixed but continuously increase over time. Most likely internal competition is the main responsible of the progressive loss of clonal heterogeneity in later stages of progression, a time frame in which complete loss of clones is always less probable.

Observing same pools of unprogressed cells undergoing glioma progression in parallel ways highlighted the stochastic nature of tumoral progression, thus uncovering that partial predetermination is possible even in first steps of gliomagenesis: in fact, some clones separated few days from transformation independently became the most represented clones in different tumors. Similarly, when tumor progression is split and low grade clones give rise to higher grade tumors in distinct immunocompetent mice, this partial predetermination shows up more clearly. The concordance seen in the clonal composition of secondary tumors suggest that some clones are more progressed than others and can more easily go through final progression steps in which they qualitatively change and gain malignity rather than selectively expand.

Materials and Methods

Cellular Biology

Cell lines and culture media

Primary cultures of dissociated tumor cells were grown in Matrigel-coated flasks and using a medium designed for the culture of neural stem cells, composed by High glucose DMEM (Invitrogen), F12 (Invitrogen), 1% B27 (Invitrogen), 0.01% bFGF, 0.01% EGF (PreProTech).

NIH-3T3 cells were grown in DMEM supplemented with 10% Fetal Bovine Serum (FBS)

Cell cultures

Cells were grown in incubator at 37°C and 5% CO_2. Adherent cells grow on the bottom of the plate forming a monolayer and, when they reached confluence, they were detached using Tripsin-EDTA (Gibco), which was then neutralized by DMEM containing FBS 10%. Neutralized medium was then removed by centrifugation and cells were suspended in their medium and plated.

Retroviral particles production

Replication-defective retroviral supernatants were prepared by transiently transfecting plasmid libraries into Phoenix packaging cells, which constitutively express the necessary retroviral proteins and when transfected with constructs containing retroviral LTR (Long Terminal Repeat) and Ψ sequences they are able to encapsulate them in retroviral particles [73].

Cell supernatant was collected after 48 and 72h after transfection and the particles were concentrated by centrifugation (2h at 15000g). Viral preparation was then aliquoted and stored at -80°C.

To determine the viral titer, low amount of virus particles was used to infect NIH-3T3 cells, plated on coverslips. 48h after infection, reporter-positive clones were counted under epifluorescence microscope.

Molecular Biology

DNA digestion using Restriction Enzymes

Restriction enzymes were used for molecular cloning and to validate plasmid sequences. In every case digestion were carried 2h at optimal enzyme temperature using a thermoblock. The amount of used enzyme is 3 units/µg of DNA. Restriction Enzymes were purchased by New England Biolabs or Promega.

Fragment purification or visualization were performed by agarose gel electrophoresis, using 1% or 2% agarose. Agarose Gels were analyzed under UV or blue light and, in case of fragment purification, bands were excised and extracted using Qiagen Gel Extraction kit, using manufacturer's protocol.

Subcloning efficiency competent cells Transformation

Non-heterogeneous vectors and small-scale tests were amplified by subcloning efficiency *E. coli* cells (Thermo Fisher). Cells were transformed by heat-shock by keeping them 45'' at 42°C after being chilled on ice for 30'. After heat shock cells were diluted in SOC medium and pre-inoculated 45' at 37°C.

After then, cells were plated on agar plates containing selective antibiotic and kept growing at 37°C overnight.

Plasmid library generation

Double-stranded degenerated oligonucleotides generation

Degenerated oligonucleotides mixtures have been synthetized as single strands by TIB molBiol. To obtain double stranded mixtures, oligos were incubated 40' at 72°C with Taq Polymerase (Promega) and an oligonucleotide complementary to the non-degenerated 3'end of oligos. The reaction product was isolated in Agarose gel and used for Gibson Assembly.

Gibson Assembly

400 ng of linearized vector (about $4x10^{10}$ molecules) and degenerated oligos in 1:25 ratio were incubated 1h at 60°C in a reaction mix containing dNTPs, 3' DNA exonuclease, DNA polymerase and Taq ligase (HiFi DNA assembly kit, New England Biolabs)

High efficiency electrocompetent cells electroporation

Electrocompetent cells were transformed using GenePulser (BioRad) set at 2.0 kV, 200 Ω and 25 µF. 100 ng of plasmide (about 10^{10} molecules) were used.

After electroporation cells have been diluted in SOC medium and pre-inoculated 1h at 37°C. Part of them was plated to verify transformation efficiency and the remaining part has been diluted in 1l and kept growing overnight until reaching 1 OD of density.

Estraction of plasmid DNA

Plasmid DNA was extracted from cells using Nucleobond extraction kit (Macherey-Nagel). For the entire library preparation, 3 columns were used under manufacturer's indications. After precipitation DNA was washed using 70% ethanol and resuspended in 300 µl of nuclease-free water.

Animal procedures

All animal procedures were approved by the internal committee for the protection of animals used for scientific purposes (OPBA) of the Ospedale policlinico San Martino and by the Italian

Ministry of Health according to the Italian Law D. lgs 26/2014 and the European Directive 2010/63/EU of the European Parliament. In all the experiments were used the C57BL/6J and the NOD.CB17-Prkdc^{scid}/J strains (hereinafter respectively referred as C57BL/6 and NOD/SCID) from Jackson Laboratory.

For all surgery procedures, mice were anesthetized by a drug cocktail containing Metedomine 0.5mg/Kg, Midazolam 5mg/kg and Fentanyl 0.005mg/kg, and awakened by mixture of Atipamezol 2.5mg/kg, Flumacenil 0.5mg/kg and Naloxon 1.2mg/kg.

Intraventricolar injections

In utero intraventricular injections were performed on anesthetized pregnant dams at the fourteenth day of gestation. Females were anesthetized and following laparotomy, uterine horns were exposed. Embryos were injected using a glass capillary needle within the telencephalic ventricles with approximately 2 µl of retroviral suspension containing 1% polybrene (Sigma) to facilitate cell infection and Fast green (Sigma) as contrast agent. After birth, injected animals were daily monitored and sacrificed at the appearance of neurologic symptoms.

Resolbable suture was used before awakening the animals.

Intracranic injection of tumoral cells

Tumor cells were injected in deeply anesthetized adult C57/bl6 or NOD/SCID mice mounted on a stereotaxic apparatus. Up to 5 µl of cell suspension for each mouse, containing 30000 to 50000 acutely dissociated cells, were injected using a Hamilton syringe (Bregma coordinates: AP, 1.0 mm; L, 1.5 mm left and 2.5 mm below the skull surface). Animals were then monitored daily for the onset of neurologic symptoms.

Dissection and dissociation of tumors

Brains were explanted from mice and positioned under fluorescence stereomicroscope (Leica): tumor areas are recognizable by GFP expression, hyper vascularization and necrosis.

Tumor and surrounding areas were dissected and enzymatically digested with trypsin for 20' at 37°C and then mechanically dissociated using a Pasteur pipette. If the percentage of GFP positive cells in the resulting population was less than 10%, cells were enriched by FACS sorting using a high recovery mask.

NGS sequencing

Ion Torrent Sequencing

Plasmid or genomic DNA samples have been amplified by PCR using primers complementary to barcode flanking regions, and containing IonA and IonP1 adaptor sequences for Ion Torrent platform. PCRs were performed in a real-time thermocycler (Bio-Rad) in order to observe the reaction and stop it in exponential phase to avoid drifts that could lead to non-representative sampling. Amplicons have been purified from primers by agarose gel electrophoresis and purified using Qiagen Gel Extraction Kit.

Sequencing libraries have been sequenced using Ion PGM machine (Thermo Fischer) and 314 chips.

Illumina Sequencing

Genomic DNA samples have been amplified by PCR using primers complementary to barcode flanking regions, containing the first part of Illumina adaptor sequences and a sample-specific index sequence 6 nt long. Each sample was then identified by the couple of indexes present in the forward and in the reverse primer. Using 10 different forward primers and 8 different reverse primers I had the possibility to pool up to 80 samples in a single sequencing run.

After the first PCR reaction, performed in a Bio-Rad real-time thermocycler and stopped in the exponential phase, each PCR reaction was individually diluted 1:30 and used as template for a second short PCR reaction, using primers containing the second part of Illumina adaptors and a second couple of sample-specific Indexes, chosen to coherently correspond to the first couple. In this way it is possible to recognize and discard almost any index-hopping-mediated mislabeling.

Data Analysis

Each sequencing was analyzed using "Biostrings" library of Bioconductor R package. Self-developed scripts were used to align each read from fastQ files to the retroviral backbone region, and using the alignment coordinate it was possible to retrieve the barcode sequence from each single read. When present, the sample specific indexes were also retrieved and used to couple each barcode with its sample.

References

1. Poppleton, H. and R.J. Gilbertson, *Stem cells of ependymoma.* Br J Cancer, 2007. **96**(1): p. 6-10.
2. Schwartzbaum, J.A., et al., *Epidemiology and molecular pathology of glioma.* Nat Clin Pract Neurol, 2006. **2**(9): p. 494-503; quiz 1 p following 516.
3. Pardridge, W.M., *Drug transport across the blood-brain barrier.* J Cereb Blood Flow Metab, 2012. **32**(11): p. 1959-72.
4. Lun, M., et al., *The natural history of extracranial metastasis from glioblastoma multiforme.* J Neurooncol, 2011. **105**(2): p. 261-73.
5. Sun, B.L., et al., *Lymphatic drainage system of the brain: A novel target for intervention of neurological diseases.* Prog Neurobiol, 2018. **163-164**: p. 118-143.
6. Pansera, F. and E. Pansera, *An explanation for the rarity of extraaxial metastases in brain tumors.* Med Hypotheses, 1992. **39**(1): p. 88-9.
7. Giese, A., et al., *Substrates for astrocytoma invasion.* Neurosurgery, 1995. **37**(2): p. 294-301; discussion 301-2.
8. Fidler, I.J., *Tumor heterogeneity and the biology of cancer invasion and metastasis.* Cancer Res, 1978. **38**(9): p. 2651-60.
9. Davidson, N.E., et al., *AACR Cancer Progress Report 2016.* Clinical Cancer Research, 2016. **22**(19 Supplement): p. S1-S137.
10. Friedmann-Morvinski, D., *Glioblastoma heterogeneity and cancer cell plasticity.* Crit Rev Oncog, 2014. **19**(5): p. 327-36.
11. Zulch, K.J., *Principles of the new World Health Organization (WHO) classification of brain tumors.* Neuroradiology, 1980. **19**(2): p. 59-66.
12. Jaffe, E.S., et al., *Classification of lymphoid neoplasms: the microscope as a tool for disease discovery.* Blood, 2008. **112**(12): p. 4384-99.
13. Jaffe, E.S., *The Microscope as a Tool for Disease Discovery-A Personal Voyage.* Annu Rev Pathol, 2017. **12**: p. 1-24.
14. Visvader, J.E., *Cells of origin in cancer.* Nature, 2011. **469**(7330): p. 314-22.
15. Garraway, L.A. and W.R. Sellers, *Lineage dependency and lineage-survival oncogenes in human cancer.* Nat Rev Cancer, 2006. **6**(8): p. 593-602.
16. Husby, S. and K. Gronbaek, *Mature lymphoid malignancies: origin, stem cells, and chronicity.* Blood Adv, 2017. **1**(25): p. 2444-2455.
17. Krivtsov, A.V., et al., *Cell of origin determines clinically relevant subtypes of MLL-rearranged AML.* Leukemia, 2012. **27**: p. 852.
18. Jellinger, K., *Glioblastoma multiforme: morphology and biology.* Acta Neurochir (Wien), 1978. **42**(1-2): p. 5-32.

19. Alcantara Llaguno, S.R. and L.F. Parada, *Cell of origin of glioma: biological and clinical implications.* Br J Cancer, 2016. **115**(12): p. 1445-1450.
20. Holland, E.C., et al., *Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice.* Nat Genet, 2000. **25**(1): p. 55-7.
21. Alcantara Llaguno, S., et al., *Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model.* Cancer cell, 2009. **15**(1): p. 45-56.
22. Uhrbom, L., et al., *Ink4a-Arf loss cooperates with KRas activation in astrocytes and neural progenitors to generate glioblastomas of various morphologies depending on activated Akt.* Cancer Res, 2002. **62**(19): p. 5551-8.
23. Merkle, F.T., Z. Mirzadeh, and A. Alvarez-Buylla, *Mosaic organization of neural stem cells in the adult brain.* Science, 2007. **317**(5836): p. 381-4.
24. Sharma, M.K., et al., *Distinct genetic signatures among pilocytic astrocytomas relate to their brain region origin.* Cancer Res, 2007. **67**(3): p. 890-900.
25. Taylor, M.D., et al., *Radial glia cells are candidate stem cells of ependymoma.* Cancer Cell, 2005. **8**(4): p. 323-35.
26. Sementino, E., et al., *Inactivation of Tp53 and Pten drives rapid development of pleural and peritoneal malignant mesotheliomas.* J Cell Physiol, 2018. **233**(11): p. 8952-8961.
27. Ahronian, L.G. and B.C. Lewis, *Using the RCAS-TVA system to model human cancer in mice.* Cold Spring Harb Protoc, 2014. **2014**(11): p. 1128-35.
28. Hambardzumyan, D., et al., *The probable cell of origin of NF1- and PDGF-driven glioblastomas.* PloS one, 2011. **6**(9): p. e24454-e24454.
29. Lindberg, N., et al., *Oligodendrocyte progenitor cells can act as cell of origin for experimental glioma.* Oncogene, 2009. **28**(23): p. 2266-75.
30. Persson, A.I., et al., *Non-stem cell origin for oligodendroglioma.* Cancer cell, 2010. **18**(6): p. 669-682.
31. Liu, C., et al., *Mosaic analysis with double markers reveals tumor cell of origin in glioma.* Cell, 2011. **146**(2): p. 209-221.
32. Alcantara Llaguno, S.R., et al., *Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes.* Cancer cell, 2015. **28**(4): p. 429-440.
33. Jiang, Y., et al., *Glioblastoma Cell Malignancy and Drug Sensitivity Are Affected by the Cell of Origin.* Cell Rep, 2017. **18**(4): p. 977-990.
34. Holland, E.C., et al., *Astrocytes give rise to oligodendrogliomas and astrocytomas after gene transfer of polyoma virus middle T antigen in vivo.* The American journal of pathology, 2000. **157**(3): p. 1031-1037.
35. Lindberg, N., et al., *Oncogenic signaling is dominant to cell of origin and dictates astrocytic or oligodendroglial tumor development from oligodendrocyte precursor cells.* J Neurosci, 2014. **34**(44): p. 14644-51.
36. Ceccarelli, M., et al., *Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma.* Cell, 2016. **164**(3): p. 550-563.
37. Brennan, C., et al., *Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations.* PLoS One, 2009. **4**(11): p. e7752.
38. Phillips, H.S., 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): p. 157-73.
39. 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, 2010. **17**(1): p. 98-110.

40. Fidler, I.J. and I.R. Hart, *Biological diversity in metastatic neoplasms: origins and implications.* Science, 1982. **217**(4564): p. 998-1003.
41. Farabegoli, F., et al., *Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH.* Cytometry, 2001. **46**(1): p. 50-6.
42. Park, S.Y., et al., *Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype.* J Clin Invest, 2010. **120**(2): p. 636-44.
43. Snuderl, M., et al., *Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma.* Cancer Cell, 2011. **20**(6): p. 810-7.
44. Soeda, A., et al., *The evidence of glioblastoma heterogeneity.* Sci Rep, 2015. **5**: p. 7979.
45. 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, 2012. **109**(8): p. 3041-6.
46. Sottoriva, A., et al., *Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics.* Proc Natl Acad Sci U S A, 2013. **110**(10): p. 4009-14.
47. Nowell, P.C., *The clonal evolution of tumor cell populations.* Science, 1976. **194**(4260): p. 23-8.
48. Gao, R., et al., *Punctuated copy number evolution and clonal stasis in triple-negative breast cancer.* Nat Genet, 2016. **48**(10): p. 1119-30.
49. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-4.
50. Navin, N., et al., *Inferring tumor progression from genomic heterogeneity.* Genome Res, 2010. **20**(1): p. 68-80.
51. Hebert, P.D., et al., *Biological identifications through DNA barcodes.* Proc Biol Sci, 2003. **270**(1512): p. 313-21.
52. Tan, S.L., et al., *DNA barcoding herbaceous and woody plant species at a subalpine forest dynamics plot in Southwest China.* Ecol Evol, 2018. **8**(14): p. 7195-7205.
53. Janda, J.M. and S.L. Abbott, *16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls.* J Clin Microbiol, 2007. **45**(9): p. 2761-4.
54. Sollai, M., et al., *A combined lipidomic and 16S rRNA gene amplicon sequencing approach reveals archaeal sources of intact polar lipids in the stratified Black Sea water column.* Geobiology, 2018.
55. Jovel, J., et al., *Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics.* Front Microbiol, 2016. **7**: p. 459.
56. Schepers, K., et al., *Dissecting T cell lineage relationships by cellular barcoding.* J Exp Med, 2008. **205**(10): p. 2309-18.
57. Gerlach, C., et al., *One naive T cell, multiple fates in CD8+ T cell differentiation.* J Exp Med, 2010. **207**(6): p. 1235-46.
58. Levy, S.E. and R.M. Myers, *Advancements in Next-Generation Sequencing.* Annu Rev Genomics Hum Genet, 2016. **17**: p. 95-115.
59. Blundell, J.R. and S.F. Levy, *Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer.* Genomics, 2014. **104**(6 Pt A): p. 417-30.
60. Gerrits, A., et al., *Cellular barcoding tool for clonal analysis in the hematopoietic system.* Blood, 2010. **115**(13): p. 2610-8.
61. Lu, R., et al., *Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding.* Nat Biotechnol, 2011. **29**(10): p. 928-33.
62. Bhang, H.E., et al., *Studying clonal dynamics in response to cancer therapy using high-complexity barcoding.* Nat Med, 2015. **21**(5): p. 440-8.
63. Giessler, K.M., et al., *Genetic subclone architecture of tumor clone-initiating cells in colorectal cancer.* J Exp Med, 2017. **214**(7): p. 2073-2088.

64. Hata, A.N., et al., *Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition.* Nat Med, 2016. **22**(3): p. 262-9.
65. Lan, X., et al., *Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy.* Nature, 2017. **549**(7671): p. 227-232.
66. Kalhor, R., P. Mali, and G.M. Church, *Rapidly evolving homing CRISPR barcodes.* Nat Methods, 2017. **14**(2): p. 195-200.
67. McKenna, A., et al., *Whole-organism lineage tracing by combinatorial and cumulative genome editing.* Science, 2016. **353**(6298): p. aaf7907.
68. Livet, J., et al., *Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system.* Nature, 2007. **450**(7166): p. 56-62.
69. Weber, T.S., et al., *Site-specific recombinatorics: in situ cellular barcoding with the Cre Lox system.* BMC Syst Biol, 2016. **10**(1): p. 43.
70. Peikon, I.D., D.I. Gizatullina, and A.M. Zador, *In vivo generation of DNA sequence diversity for cellular barcoding.* Nucleic Acids Res, 2014. **42**(16): p. e127.
71. Calzolari, F., et al., *Tumor progression and oncogene addiction in a PDGF-B-induced model of gliomagenesis.* Neoplasia, 2008. **10**(12): p. 1373-82, following 1382.
72. Appolloni, I., et al., *Progression from low-to high-grade in a glioblastoma model reveals the pivotal role of immunoediting.* Cancer Lett, 2018.
73. Pear, W.S., et al., *Production of high-titer helper-free retroviruses by transient transfection.* Proc Natl Acad Sci U S A, 1993. **90**(18): p. 8392-6.