Evolutionary dynamics of residual disease in human glioblastoma

*Inmaculada Spiteri1,\*, Giulio Caravagna1,\*, George Cresswell1, Alexandra Vatsiou1, Daniel Nichol1, Ahmet Acar1, Luca Ermini1, Kate Chkhaidze1, Benjamin Werner1, Anestis Touloumis2, Enrico Brognaro3, , Guido Sanguinetti5, Sara G.M. Piccirillo4,§, Colin Watts6,§, Andrea Sottoriva1,§.*

1. Evolutionary Genomics & Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK.

2. School of Computing, Engineering and Mathematics, University of Brighton, Brighton, UK

3.

4. Department of Internal Medicine, Division of Hematology and Oncology, University of Texas Southwestern Medical Center, Dallas, USA

5. School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK.

6. Birmingham Brain Cancer Program, Institute of Cancer Genome Sciences, University of Birmingham, Birmingham, UK

\* These authors contributed equally to this work

*§* Correspondence should be addressed to:

sara.piccirillo@utsouthwestern.edu

[c.watts.2@bham.ac.uk](mailto:c.watts.2@bham.ac.uk)

[andrea.sottoriva@icr.ac.uk](mailto:andrea.sottoriva@icr.ac.uk)

# Abstract (200-300)

Glioblastoma is the most common and aggressive adult brain malignancy against which conventional surgery and chemoradiation provide limited benefit. Even when complete surgical resection is achieved, recurrence may occur distally, driven by cancer clones that are often genomically distinct from those in the primary tumour. Glioblastoma cells display a characteristic infiltrative phenotype, invading the surrounding tissue and often spreading across the whole brain. We hypothesised that cancer cells responsible for relapse reside in two compartments of residual disease that are left behind after resection and treatment: the infiltrated normal brain parenchyma, and the sub-ventricular zone (SVZ). In this proposed model, residual disease subclones diverge early during glioblastoma evolution, may remain dormant in the normal parenchyma, and are responsible for recurrence. To test this, we performed whole-exome sequencing of 69 multi-region samples collected using fluorescence-guided resection from 11 patients, including the infiltrating tumour margin (M) and the SVZ for each case, as well as matched blood. We used a phylogenomic approach to dissect the spatio-temporal evolution of each tumour and unveil the relation between residual disease and the main tumour mass. We also analysed two cases of primary-recurrence samples with matched residual disease. Our results suggest that the infiltrative phenotype is an early evolutionary trait of glioblastoma, giving rise to the tumour mass detected at surgery. After treatment, the same infiltrative clone may trigger the growth of a recurrent tumour, thus representing the ‘missing link’ between the primary tumour and the recurrence. These results are consistent with recognised disease aetiology and prognosis and suggest that targeting cells in the infiltrated brain parenchyma may improve patient’s outcome.

# Introduction

Glioblastoma (GB) is a lethal brain cancer against which effective therapeutic options are lacking [1]. The disease is characterised by variegated clinical phenotypes [2-4] and intra-tumour heterogeneity (ITH) [5-7]. The disease aetiology and clinical course have distinct features compared to other cancers. Unlike other solid tumours, glioblastoma rarely metastasises outside the brain, but it invariably recurs, limiting the median survival to approximately 14 months[1]. In approximately 80% of aggressively treated patients, disease progression/recurrence occurs within 2cm of the resection margin. In the remaining patients, even when complete surgical removal of the primary lesion is possible, the tumour recurs distally [6-8] and even drastic hemispherectomy procedures fail to eradicate the disease [9]. Cancer cells from these distal recurrent lesions, despite sharing a common ancestor with the primary tumour, are often genomically distinct [6-8]. Moreover, at diagnosis glioblastoma already displays a characteristic infiltrative phenotype, invading the surrounding brain tissue and often diffusely infiltrating the whole brain [10]. We have previously shown that malignant clones present in the sub-ventricular zone (SVZ), a known neural stem cell germinal niche, represent either tumour precursor cells or an infiltrating population [11]. Indeed, infiltration is ubiquitous in glioblastoma, with cells migrating through diverse regions of the brain microenvironment including white matter tracts [12] and blood vessels (ref. Calabrese C, Cancer Cell 2007). In addition, up to 10% of glioblastoma cases present as multifocal disease at diagnosis [13], a rare occurrence in other solid tumours.

The accumulating clinical and genomic evidence points at a different model of tumour initiation and recurrence in glioblastoma with respect to the classical step-wise transformation and diversification of malignant phenotypes, which posits that solid tumours first grow locally, then invade the surrounding tissues and only finally spread.

Here, we propose that after glioblastoma initiation (Figure 1A), due to its inherent infiltrative phenotype, cancer cells quickly spread and colonise the brain parenchyma & the SVZ (Figure 1B). Although infiltrative cells are initially characterised by diffuse migration and low proliferation, different microenvironments and external changes - may drive proliferation and ultimately the formation of a large mass (doi: 10.18632/oncotarget.7454. & Figure 1C). After treatment, infiltrative cells in the brain parenchyma and in the SVZ can drive relapse. Thus, even in the case of optimal treatment, when the tumour mass has been completely removed (Figure 1D,E), we predict that residual disease in M and SVZ will trigger new growth, giving rise to distal recurrence (Figure 1F). We argue that the recent seminal studies performed on primary-recurrent matched glioblastoma samples [6-8] point at the infiltrative population as the “missing link”, connecting the primary and the recurrent malignant clone in the evolution of the disease. Hence, considering this body of biological, clinical and genomic evidence, and contrary to the other solid malignancies, the infiltrative clone(s) may originate *before* the proliferative clone(s) in glioblastoma.

# Materials and Methods

**Patient cohort and samples**

Sixty-nine tissue samples were collected from neurosurgical fluorescence-guided resections performed on eleven GBM patients. Between 5 and 9 multiple samples from the tumor (T), sub-ventricular zone (SVZ) and margin (M) areas that were at least 1cm apart were collected from each patient. In the case of the two primary/recurrence cases, 3 samples (T, SVZ, M) were taken during the primary and secondary surgical resections to a total of 6 specimens per patient. Thirty 10-um cryosections were taken from each frozen tissue for DNA extraction using the DNeasy Blood & Tissue kit (Qiagen). Patient informed consent was obtained and tissue collection/storage protocols were compliant with the UK Human Tissue Act 2004 and approved by the Local Regional Ethics Committee (LREC ref 04/Q0108/60). No difference in 5-ALA labelling capacity was observed between patients.

**Whole-exome and targeted sequencing**

Between 100-300ng of DNA from each of the 69 tumour specimens and 11 blood samples were sent to GATC Biotech (Germany) for whole exome sequencing using the Agilent SureSelectXT Human All Exon V5 Kit. Exome sequencing cohort had a median coverage of 157X (min. 108X, max. 187X). Further, a custom panel for 891 SNVs (covered by a total of 5,090 amplicons) identified from the exome sequencing data was designed using Agilent’s Haloplex technology. In addition, we designed a separate Agilent SureSelect XT2 capture panel to specifically validate 1,054 SNVs found in the M and SVZ samples. Both amplicon and targeted capture libraries were sequenced on an Illumina HiSeq2500 obtaining a median coverage of 4,050X (min. 8X, max. 31,349X) and 1,154X (min. 5X, max. 4,351X) respectively, in called variants.

**Bioinformatics Analysis**

Purity estimation: Purity estimates were gathered from a potential clonal diploid mutation (PCDM) present in the tumour mass samples and estimation by ASCAT [14] derived from the copy number solution.

Copy number analysis: Sequenza was used to identify heterozygous single nucleotide polymorphisms (SNPs) in the WES (0.4-0.6 allele frequency in the matched normal sample) and normalise depth ratios for GC content [15]. Loci were filtered for a minimum of 25 reads in the matched normal sample. Log2 ratios (LRR) were derived from the depth ratios by calculating the log (base 2) of the depth ratio and subtracting by the global median. LRR outliers were smoothed using CGHcall [16]. The mirrored allele frequencies of the heterozygous loci were segmented using piece-wise constant fitting (PCF) [17]. If BAF values in segments are considered not to be drawn from a normal distribution expected in allele balance (BAF = 0.5, Kolmogorov-Smirnov test, p < 0.05), a two-component Gaussian mixture model was fitted to the B-allele frequency (BAF) values of the segment utilising mixtools version 1.0.4 [18], in order to estimate BAFs representative of the major allele. Major allele BAF and the LRR of the genome segments were used as input for ASCAT to estimate tumour purity and ploidy, limiting the minimum purity of the solution to the lower 95% binomial confidence limit (Wilson method) of a PCDM purity estimate adjusted for a tetraploid solution. Ploidy parameter space was restricted in tumours manually determined to have high ploidy states, additionally purity and ploidy was preset if a solution was not determined. The purity and ploidy of the ASCAT solution was used to assess the clonality of each segment using the Battenberg methodology

**[19]**. If a segment was considered subclonal, the copy number state with the highest prevalence was taken.

Identification and validation of somatic variants: adapter trimming was performed with Skewer v0.1.126 [20] with minimum read length after trimming 35 and mean quality value before trimming of 10. Trimmed reads were aligned to the full human reference genome hg19 with Burrows-Wheeler Aligner (BWA) v0.7.12 [21]. PCR duplicates were marked using Picard tools. Joint mutation calling between multiple samples from the same patient was performed per patient using a combination of Platypus v0.8.1 [22] with biased prior (‘source’ option) for mutations called by Mutect2

[23] on single tumour-normal pairs. This allowed us to exploit the sensitivity of Mutect2 with the joint calling capability of Platypus. The following filtering criteria were used to call somatic variants in WES samples: i) only variants with Platypus filter PASS, alleleBias, Q20, QD, SC and HapScore were kept, ii) minimum coverage and genotype quality of 10 was required iii) variants in segmental duplicated regions and centromeric regions were removed, iv) minimum of 3 reads covering the variant in at least one of the tumour samples per patient were considered, v) 0 number of reads covering the variant in the germline sample, and vi) genotype of the 0/0 in the germline sample. Only somatic alterations and indels with a Variant Allele Frequency (VAF) >5% were considered. Somatic variants were annotated both with CAVA [24] and VEP [25]. SNV calling on the targeted capture samples was performed using Platypus in genotyping mode. Somatic SNVs with minimum genotype quality of 10, minimum coverage of 300 and identified by a minimum of 10 reads were considered for further analysis. Somatic mutations that failed the validation in all samples per patient were removed otherwise VAF is indicated as NA in the failed sample. All somatic mutation calls from TES and WES panels are available in VCF files as Supplementary Data.

Driver genes: the complete set of SNVs was compared to a list of known putative driver genes in glioblastoma from ref [8].

Cancer cell fraction estimation: Cancer cell fraction of each variant is calculated using the VAF, estimated purity of the sample and total copy number state of the segment in which the variant belongs [26]. The estimated purity calculated by ASCAT was used unless the estimation was 1, then the PCDM purity estimate was used. To avoid overcalling subclonality, the number of mutated alleles was assumed to be 1.

Phylogenetic reconstruction: for each patient, SNVs identified from the whole exome sequencing panel were used to construct sample phylogenies using PAUP\* (version 4.0a) with the maximum-parsimony criterion [27]. CCF values for the SNVs were first dicotomised to produce tables indicating the presence/absence of each mutation in each sample by defining a mutation as present where CCF≥0.2. For patient 54, the SVZ sample was included in the phylogenetic analysis with mutations identified by the TES2 panel and dicotomised as above. For all patients, we considered only mutations for which the presence/absence in each sample was determined. Where a mutation was absent from a margin sample (M) for a patient, but not identified as a true negative with high probability, this mutation was omitted from the phylogenetic reconstruction. The maximum-parsimony trees were identified via a heuristic search with default settings. For each phylogenetic tree, 1000 replicates of bootstrap analysis were carried out to assess the support (Figure S**?**). Trees were rendered with the R package phytools [28](R version 3.3.2, phytools version 0.6-44).

# Statistical testing for mutations in the margin

We developed a statistical test to compute p-values for the null hypothesis that clonal mutations are missing from the *margin* samples (labelled as “M”). The test is divided in two steps: 1) *training*, in which we identify suitable clonal mutations to train a statistical model of expected read counts from whole-exome (WES) data, and 2) *test,* in which we use the trained model to compute the likelihood for a null hypothesis built from two targeted deep sequencing panels (labelled as “TES1” and “TES2”). The tests are carried out independently on each one of the input samples of a patient and corrected for multiple testing. Rejecting provides evidence that mutations missing in the margin samples but present in the primary tumour, are unlikely clonal. Combined with our phylogenetic analysis, this provides solid evidence that the margin is an ancestor of the primary tumour.

## Training a statistical model from read-counts

For each patient, we used its CCFs and CNA values to identify putative *clonal* SNVs in exomic regions. From Platypus VCF files, we selected only entrieswith:

1. A single-nucleotide mutation;
2. CCF >0.8 in *all* primary samples;
3. the *same CNA status* across all primary samples.

Discarded indel mutations are reported in Supplementary Table XXXX. Mutations that do not fulfil these conditions are likely subclonal, or are not SNVs, and neglected by this analysis. Read counts for clonal SNVs are extracted from VCF files reporting Number of Variants (NV, number of reads with the alternative allele) and Reads (NR, i.e., coverage).

We corrected read counts for CNA statusand tumour purity, before using them to train a statistical model for the test. Correction is carried out to estimate how many of the *observed reads* come from the actual tumour. The correction is a standard procedure which uses tumour purity (here estimated from WES data) and tumour copy number status *c* as follows

Factor two in the correction is the diploid copy number status of normal cells. After correction, is the number of reads, out of *r*, that are estimated coming from tumour cells. To be conservative, we assume the mutation multiplicity (number of allele copies that bear the mutation) is always 1.

With the corrected coverage value, we can fit a *Beta-Binomial* with number of trials , and successes NV. A success in this experiment is the detection of a read with the mutant allele; the overall Binomial sampling consists in the repetition of this experiment with reads (i.e., the number of reads at a locus). The contribution of the Beta distribution is to capture uncertainty over the success probability of the experiment. Hence, this statistical model describes the probability of observing mutated at coverage as a function of the Beta distribution

This quantity is related to the so-called Variant Allele Frequency (VAF), which is defined to be . This compound model extends Binomial sampling with over-dispersion effects that better account for non-uniform coverage in sequencing assays [CITE.WHO?]; technically, it is a Binomial distribution whose parameter follows a Beta distribution with hyper-parameters and , greater than 0.

Multiple Beta-Binomial model was trained for each copy number status for the input SNVs, and each WES sample of the primary tumour. By separating read counts by copy number status, we can adjust data for every non-diploid SNV in a more precise and consistent way. To train a model at minimum acceptable quality for the next test, we discarded all combination of parameters for which we do not have, at least, 10 available SNVs. When that is the case, we removed from downstream analysis this configuration of copy number status and input sample.

To learn the model parameters and from data, we used the Maximum Likelihood fitting procedure vglm for *vector generalized linear models* that is implemented in the R package VGAM (version 1.0-5).

## Testing

We identified testable SNVs from two deep-sequencing targeted panels. Because panels have ~3000x coverage, we require each SNV to have at least reads with the variant allele (NV). When that is not the case, the SNV is considered missing from the panel, and it qualifies as suitable for our test. We particularly care about the ones that are missing in the margin sample (M): if the margin was ancestral to the primary tumour, we expect it to lack some SNVs that are clonal in the primary WES samples.

If a patient has more than one margin sample (e.g. Patient A23), we require the mutation to be absent across *all* margin samples. SNVs selected in this way appear indeed clonal in the primary tumour but are missing in the margin biopsy from the targeted panels. Some patients have no testable mutations (e.g. Patient 42). The SNVs that we detected from both targeted panels are pooled together, and their number of reads (NR) from the VCF files is stored; if an SNV is detected from both panels, we sum the NR values from both panels.

Read counts from the margins are corrected as with the training set. Copy number status for mutations in the margin is the same for the training set, by design choice. Purity correction instead requires some considerations. Estimation of purity for margin samples is hard because of the apparent high contamination of normal cells. Standard tools like PurBayes and Sequenza have estimated purity values below (30%) in most cases (data not shown). To be conservative, however, we have used a fixed, worst-case low purity estimate of (1% tumour, 99% normal). This value is much lower than the margin’s likely true purity, and renders the test harder since the power to reject the null decreases with coverage. Thus, a conservative purity estimate leads us to rescale observed coverage to lower values (i.e., we ``throw away’’ coverage from the targeted panels).

The set of SNVs to test is divided according to the copy number status, and matched to the trained models. Each group is tested independent against all models trained from the different primary regions. The null hypothesis for testing a group of SNVs is the probability of detecting NV(with ) mutated reads at the corrected coverage, given the parameters of the matched Beta-Binomial model

.

The p-value is hence the probability of finding less than *k* reads with the variant allele at the designed locus with coverage , given the fact that we expect for a clonal SNV the number of reads to follow a Beta-Binomial model with parameters (fit from WES).

Thus, rejecting means rejecting the hypothesis that the SNV is clonal in the margin. Combined with phylogenetic analysis, this provides strong evidence of the ancestral relation between the margin and the primary, and that these missing SNVs are real *true negatives in M.* The tests are executed at confidence level , and corrected for multiple testing with the stringent correction possible (Family-wise Error Rate, via Bonferroni).

# Results

To test our hypothesis, we performed fluorescence-guided multi-region sampling of different regions from primary tumours (T1, T2, T3, …) and also collected samples from the SVZ (n=15) and the infiltrative margin M (n=11) (Figure 1G). The latter is defined by non-fluorescent tissue beyond the fluorescent tumour mass. We previously reported that this area appears histologically as normal brain and is composed by only 10% tumor cells (Piccirillo S et al. BJC 2012). Consistent with this, in our independent set of samples the tumour mass and SVZ samples displayed high tumour content whereas the margin samples were characterised by a mean purity of X% (+/-X). Clinical and follow-up information, as well as imaging was also available (Supplementary Table 1). Samples from the tumour mass and SVZ from 7/11 patients were common to our previous studies, for which we had performed copy number profiling and gene expression alone[5, 11]. The margin samples are presented in this study for the first time.

Multi-region whole-exome sequencing (mean coverage 120X) identified extensive intra-tumour heterogeneity at the level of single nucleotide variants (SNVs) and small insertions and deletions (indels). Heterogeneity at the level of putative drivers was evident in X patients, especially in XXX genes (Figure 2A). Copy number profiles inferred from the whole-exome sequencing confirmed the heterogeneity levels reported in previous studies (Figure 2B and S1). Importantly, clonal copy number events were highly recurrent, especially chromosome 7 gain/amplification (EGFR) and chromosome 10 (PTEN) loss, corroborating the findings from large scale studies[2]. We then performed ultra-deep targeted exome sequencing (TES) of a patient-specific panel of X mutations (X mutations per patient in average), achieving a mean coverage depth of 4,850±229, 95% CI, and confirmed the mutational profiles identified with whole-exome sequencing (Figure 2A, *TES1* panel).

However, the estimated purity of the margin samples was particularly low (in the order of 5%) and this was a confounding factor in our analysis because false negative calls in the margin samples would lead to erroneous reconstruction of the tumour phylogenetic tree. To tackle this problem, we designed a second targeted sequencing panel (TES2) specifically to validate whether mutations that were present in all the other tumour samples were really absent from the infiltrative areas (indicating the margin or SVZ as an ancestral subclone). We sequence the panel to extremely high depth (X). This panel confirmed that several mutations that were truncal to the tumour mass were not present in the margin sample (Figure 3A, *TES2* panel). In addition, to further support our results, we developed a new statistical method to test whether genomic variants in targeted sequencing are true negatives (indeed the mutation is not present) or false negatives (there is no power to determine with reasonable certainty that the mutation is not present). Our method is based on the idea of first fitting a beta-binomial distribution to the cancer cell fraction (CCF) of truncal mutations in the tumour. This provides the expected variant allele frequency of a mutation in a high purity sample, such as T1-4. Then, for all truncal mutations that are not found in the low-purity sample (the margin), we test whether it is likely that with the achieved depth of coverage, we did not find any read with the variant. The method is illustrated in Figure 4, and it is generally applicable to any genomic dataset to test true negatives.

Application of our method to our dataset revealed that a considerable proportion of testable mutations (see Material and Methods for details) were indeed true negatives, indicating that they were absent in the cancer cells in the margin, even when we assumed the purity to be as low as 1%. The results of the test are reported in Figure 3A, left hand side of each heatmap. This allowed to construct a reliable phylogenetic tree for each patient that included residual disease in the SVZ and the infiltrative margin M (Figure 5).

Strikingly, residual disease samples, especially the infiltrative margin M, appeared consistently at the top of the evolutionary history of each tumour, diverging early from the rest of the tumour mass in a significant proportion of patients (Figure 5). Importantly, residual disease was pivotal in the development of the distal recurrence due to the reservoir of cancer cells in this compartment (patients XX). This analysis identifies subclones present in M and SVZ as early precursors of the subclones present in the primary tumour mass. Moreover, the short branches of M and SVZ in the phylogeny suggest that those samples are characterised by slow proliferation, consistent with our previous results on M- and SVZ-derived cell populations (Piccirillo S et al BJC 2012 and ref. 11), thus indicating a quiescent/dormant phenotype of the residual disease during the evolution and treatment of the malignancy. After the primary tumour has been treated with radio- & chemotherapy the quiescent residual subclones can trigger new growth and further clonal evolution, producing the divergence that has been observed by us and others between primary and recurrent samples in GB.

This proposed model is consistent with an early onset of tumor cell infiltration that precedes aggressive growth leading to tumor mass formation. Residual disease present in the SVZ and in the infiltrative margin is the source of relapse that is inevitable in GB patients. This model is also consistent with the high incidence of multifocal lesions and the accumulating evidence of evolutionary divergence from cancer genomic data [6, 7].

# Discussion

Here we present a new model of glioblastoma initiation, development and progression. We start from a testable hypothesis based on previous clinical, biological and genomic observations and verify the hypothesis on a novel unique dataset we have generated.

A key to understand cancer is not just exposing intra-tumour heterogeneity, the natural process that underlines clonal evolution, but also to understand such heterogeneity in a way that is clinically relevant. An important aspect of genomic ITH is that it embeds the evolutionary history of the tumour, a fundamental clinical aspect that cannot be directly studied in patients. But understanding that history may be critical in developing a rationale for combinatorial therapeutics (see https://academic.oup.com/neuro-oncology/article/20/4/472/4737220).

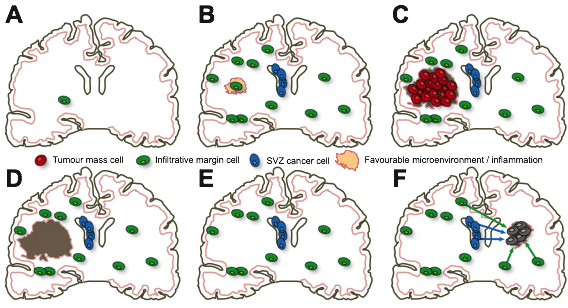
We have leveraged on the spatio-temporal decomposition of the tumour clonal architecture to understand the link between subclones in the primary tumour, in the recurrence, and in the residual disease compartment. This model implicates microenvironmental factors as an important factor in understanding GBM and the phenotypic shift of malignant cells from a migratory state to a proliferative state.

# Acknowledgments

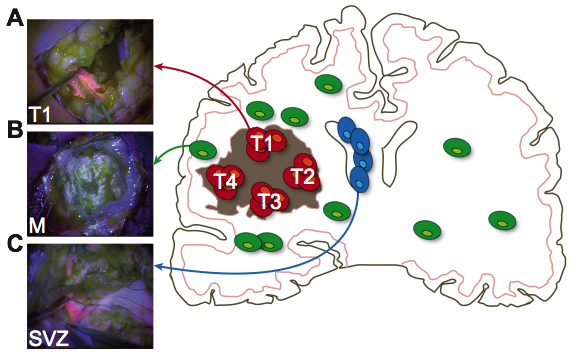
This work was supported by the Schottlander Innovation Award to the Sottoriva lab. A.S. is supported by the Wellcome Trust (202778/B/16/Z), Cancer Research UK (A22909) and the Chris Rokos Fellowship in Evolution and Cancer. We acknowledge funding from the National Institute of Health (NCI U54 CA217376) to A.S. B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. This work was also supported by a Wellcome Trust award to the Centre for Evolution and Cancer (105104/Z/14/Z).

C.W. is supported by The Brain Tumour Charity (10/136) & Cancer Research UK (C9221/A11869 & C9221/A16700). We acknowledge support from the NIHR tissue bank at the University of Cambridge. S.P. was supported in this work by The Brain Tumour Charity (10/136).

# Figure Legends



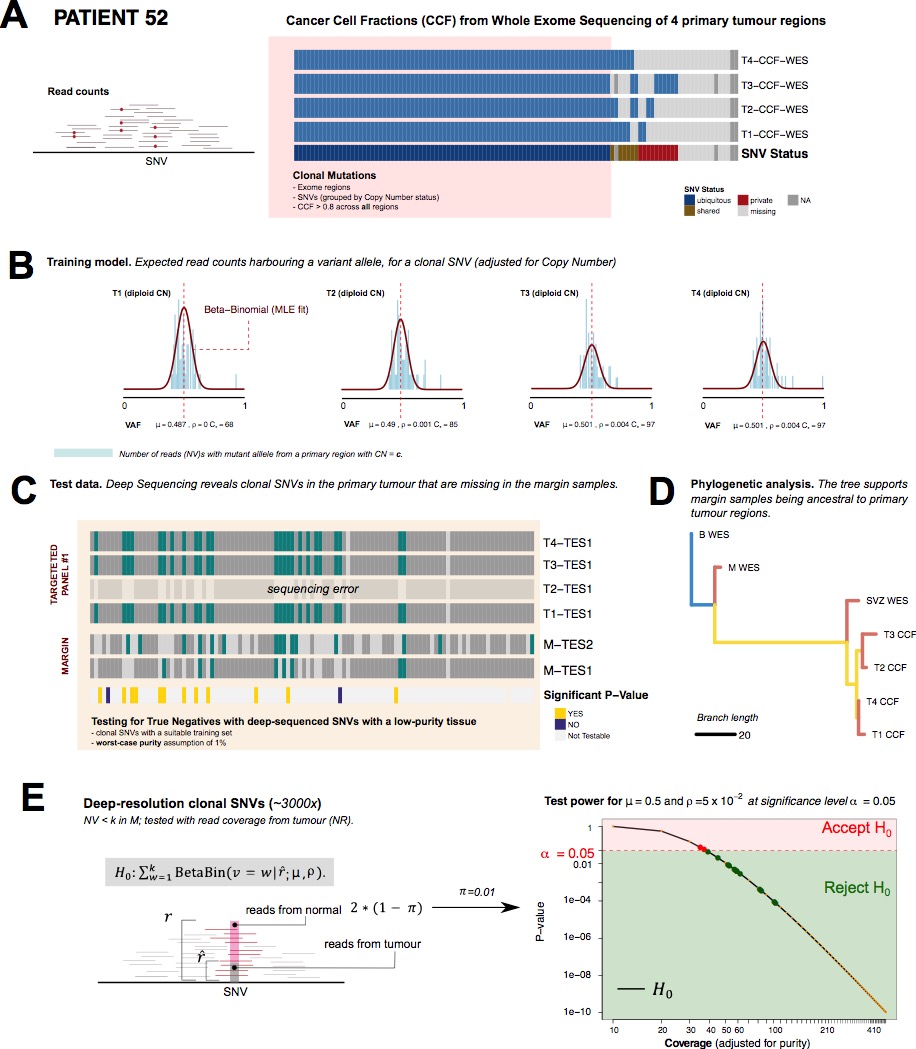
***Figure 1. Hypothesis of glioblastoma development. (A)*** *After malignant transformation, we hypothesise that glioblastoma cells extensively infiltrate the normal brain parenchyma by exploiting their intrinsic migratory phenotype along white matter tracts (green cells).* ***(B)*** *Infiltration may involve the Sub-Ventricular Zone (SVZ) a known reservoir of neural stem cells (blue cells).* ***(C)*** *When infiltrative cells encounter the right microenvironment or are triggered by a stimulus that fosters proliferation, they grow into a large mass. This model is consistent with universal relapse, extensive infiltration even in early tumours and the common incidence of multifocal disease in glioblastoma (>10%), a rare condition in other solid tumours. Hence, we hypothesise that the infiltrative clone(s) may arise before the proliferative clone(s) and not vice versa as in the large majority of solid malignancies.* ***(D)*** *At surgery, the primary tumour mass only is removed.* ***(E)*** *However, infiltrative cells in the normal brain parenchyma and SVZ are left behind.* ***(F)*** *Residual glioblastoma cells infiltrated throughout the brain then give origin to relapse tumours even at distant locations.*

******

***Figure 2. Study design and multi-region sampling. (A)*** *The large majority of patients present at diagnosis with a large tumour mass that is positive for 5-ALA fluorescence.* ***(B)*** *Extensive infiltration is also present in the surrounding normal brain but cancer cells are so sparse that do not appear fluorescent.* ***(C)*** *In a subset of patients, disease is also found in the sub-ventricular zone (SVZ), which appears fluorescent and contains malignant clones. Through surgery and chemo-radiation, it is possible to extensively remove the primary tumour but it is not feasible to completely remove the infiltrative disease, nor cancer cells in the SVZ. Those represent the majority of residual disease in glioblastoma.* ***(D)*** *To study residual disease in glioblastoma we have collected multiple regions of the same tumour, as well as residual disease in the SVZ and adjacent tumour margin during surgery and subject the samples to genomic profiling and phylogenetic analysis.*

**

***Figure 3. Multi-region genomic profiles of glioblastoma residual disease. (A)*** *We report cancer cell fractions >80% for all tumour mass samples and presence/absence of mutation in all the other samples. Putative driver events are annotated. WES=whole exome sequencing; TES1=targeted exome sequencing panel 1; TES2=targeted exome sequencing panel 2. T1…4=tumour mass sample; SVZ=sub ventricular zone; M=margin.* ***(B)*** *Digital copy number alterations are reported for each sample.*

******

***Figure 4. Testing the absence of truncal somatic mutations in the infiltrative margin with patient SP52 (examplanatory case).***

*(A)* *We detect clonal SNVs in multiple samples via WES of primary tumour regions. We selected as clonal SNVs that have CCF >0.8 and the same CNA status across all samples. (B) Using read counts from clonal SNVs, we train a Beta-Binomial model of their expected Variant Allele Frequency (VAF) of clonal mutations, for each primary sample and copy number status. This model captures, for each clonal SNV, how many reads we shall expect harbouring the variant allele. (C) We use deep-sequencing data from two targeted panels to identify which clonal SNVs are missing (or at least have less than k reads) in every margin sample. The presence of these SNVs suggests that margin samples might be ancestral to primary ones. (D) Standard phylogenetic reconstruction confirms that margin samples are ancestral to the primary regions. (E) A statistical test for SNVs identified in panel C is used to test a null hypothesis that their read counts are generated from the distribution of read counts for clonal mutations in the primary regions. Rejecting the null means that we have evidence for the fact that these SNVs are not clonal in the margin, providing further evidence that the margin is ancestral to the primary regions. The power of the test increases with higher coverage; we use a conservative setting for the test, as we estimate the actual number of reads coming from tumour cells in a worst-case purity scenario of*  (1% tumour, 99% normal). *The test is also corrected for Multiple Hypothesis Testing via Bonferroni.*

# 

***Figure 5. Phylogenetic reconstruction suggests infiltrative margin as an early ancestor.*** *Consistently across the cohort, the infiltrative margin sample appears at the top of the phylogeny, indicating it contains cancer clones that are ancestral to the rest of the tumour samples. This suggests the infiltrative clone may arise earlier than the primary tumour mass.*

# References

1. Stupp R, Mason WP, Van Den Bent MJ (2005) Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma — NEJM. New England Journal of Medicine 352:987–996.

2. McLendon R, Friedman A, Bigner D, et al (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068. doi: doi:10.1038/nature07385

3. Frattini V, Trifonov V, Chan JM, et al (2013) The integrated landscape of driver genomic alterations in glioblastoma. Nature Genetics 45:1141–1149. doi: 10.1038/ng.2734

4. Brennan CW, Verhaak RGW, McKenna A, et al (2013) The Somatic Genomic Landscape of Glioblastoma. Cell 155:462–477. doi: 10.1016/j.cell.2013.09.034

5. Sottoriva A, Spiteri I, Piccirillo SGM, et al (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci USA 110:4009–4014. doi: 10.1073/pnas.1219747110

6. Kim H, Zheng S, Amini SS, et al (2015) 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

7. Kim J, Lee I-H, Cho HJ, et al (2015) Spatiotemporal Evolution of the Primary Glioblastoma Genome. Cancer Cell 28:318–328. doi: 10.1016/j.ccell.2015.07.013

8. Wang J, Cazzato E, Ladewig E, et al (2016) Clonal evolution of glioblastoma under therapy. Nature Genetics 48:768–776. doi: 10.1038/ng.3590

9. Chaichana KL, Quinones-Hinojosa A (2014) The need to continually redefine the goals of surgery for glioblastoma. Neuro-Oncology 16:611–612. doi: 10.1093/neuonc/not326

10. Silbergeld DL, Chicoine MR (1997) Isolation and characterization of human malignant glioma cells from histologically normal brain. J Neurosurg 86:525–531. doi: 10.3171/jns.1997.86.3.0525

11. Piccirillo SGM, Spiteri I, Sottoriva A, et al (2015) Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone. Cancer Res 75:194–202. doi: 10.1158/0008-5472.CAN-13-3131

12. Claes A, Idema AJ, Wesseling P (2007) Diffuse glioma growth: a guerilla war. Acta Neuropathologica 114:443. doi: 10.1007/s00401-007-0293-7

13. Patil CG, Anthony Yi, Adam Elramsisy, et al (2012) Prognosis of patients with multifocal glioblastoma: a case-control study. J Neurosurg 117:705–711.

14. Van Loo P, Nordgard SH, Lingjærde OC, et al (2010) Allele-specific copy number analysis of tumors. Proc Natl Acad Sci USA 107:16910–16915. doi: 10.1073/pnas.1009843107

15. Favero F, Joshi T, Marquard AM, et al (2014) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Annals of Oncology 26:64–70. doi: 10.1093/annonc/mdu479

16. van de Wiel MA, Kim KI, Vosse SJ, et al (2007) CGHcall: calling aberrations for array CGH tumor profiles. Bioinformatics 23:892–894. doi: 10.1093/bioinformatics/btm030

17. Nilsen G, Liestøl K, Van Loo P, et al (2012) Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics 2012 13:1 13:591. doi: 10.1186/1471-2164-13-591

18. Benaglia T, Chauveau D, Hunter DR, Young D (2009) mixtools: An RPackage for Analyzing Finite Mixture Models. J Stat Soft. doi: 10.18637/jss.v032.i06

19. Nik-Zainal S, Van Loo P, Wedge DC, et al (2012) The life history of 21 breast cancers. Cell 149:994–1007. doi: 10.1016/j.cell.2012.04.023

20. Jiang H, Lei R, Ding S-W, Zhu S (2014) Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182. doi: 10.1186/1471-2105-15-182

21. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi: 10.1093/bioinformatics/btp324

22. Rimmer A, Phan H, Mathieson I, et al (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nature Genetics 46:912–918. doi: 10.1038/ng.3036

23. Cibulskis K, Lawrence MS, Carter SL, et al (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology 31:213–219. doi: 10.1038/nbt.2514

24. Münz M, Ruark E, Renwick A, et al (2015) CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. Genome Med 7:76. doi: 10.1186/s13073-015-0195-6

25. McLaren W, Gil L, Hunt SE, Riat HS (2016) The Ensembl Variant Effect Predictor. Genome …

26. Jiang Y, Qiu Y, Minn AJ, Zhang NR (2016) Assessing intratumor heterogeneity and tracking longitudinal and spatial clonal evolutionary history by next-generation sequencing. PNAS 201522203. doi: 10.1073/pnas.1522203113

27. Swofford DL (2005) PAUP\*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta. In: Sinauer, Sunderland, MA. http://www.sinauer.com/paup-phylogenetic-analysis-using-parsimony-and-other-methods-4-0-beta.html. Accessed 9 Mar 2015

28. Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x