

Stem Cell Reports, Volume 4

Supplemental Information

Genetic and Functional Diversity of Propagating Cells in Glioblastoma

**Sara G.M. Piccirillo, Sue Colman, Nicola E. Potter, Frederik W. van Delft, Suzanne Lillis,
Maria-Jose Carnicer, Lyndal Kearney, Colin Watts, and Mel Greaves**

SUPPLEMENTARY FIGURES

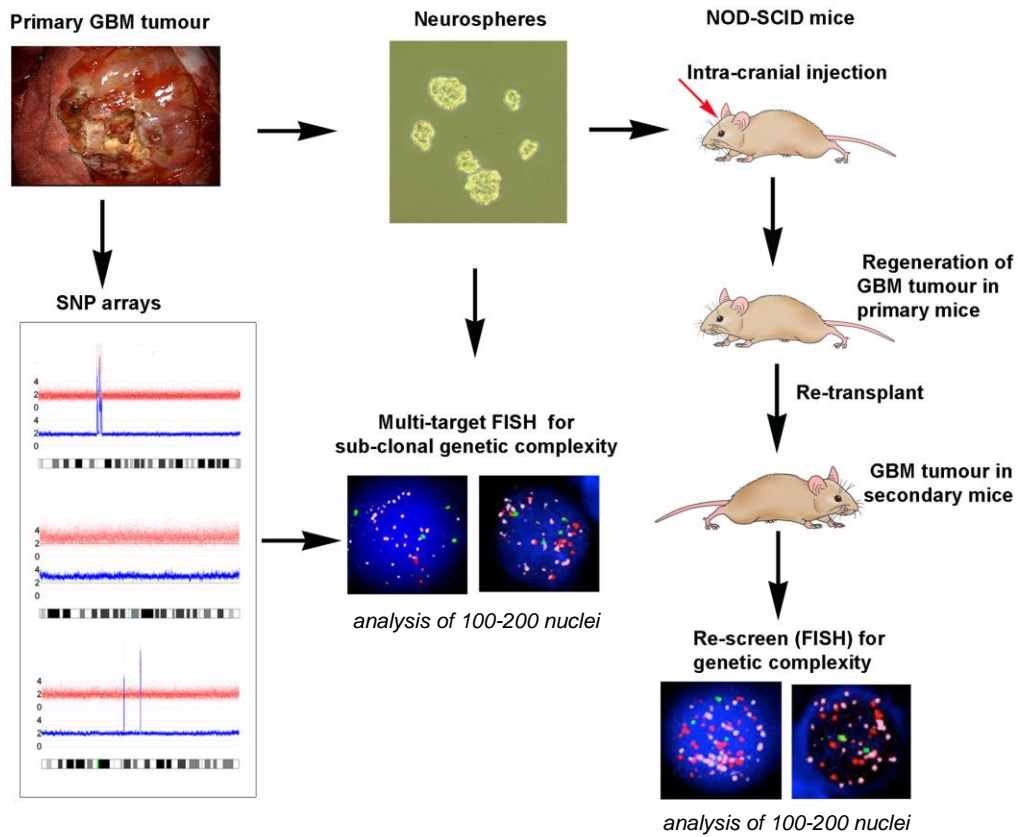


Figure S1, related to Figure 1-3. Schematic overview of the study. Primary GBM tumour was used to extract DNA for high resolution SNP (SNP6) array analysis and also to establish stable neurosphere cell cultures. High-resolution (SNP 6) arrays were performed on DNA extracted from the primary GBM tumour to identify 'driver' CNA, defined as recurrent regions of amplification or deletion. FISH probes were designed for these regions and 3 colour FISH carried out to neurosphere cells (dissociated to single cells and prepared by standard cytogenetic protocols). Neurospheres were dissociated into single cells and used for intra-cerebral transplantation (and re-transplantation) into NOD-SCID mice. FISH was carried out to the reconstituted tumours in secondary transplanted mice using the same probes used for the original FISH screen of the neurospheres. We then compared the sub-clonal genetic structure in the neurospheres and tumours generated in the mice.

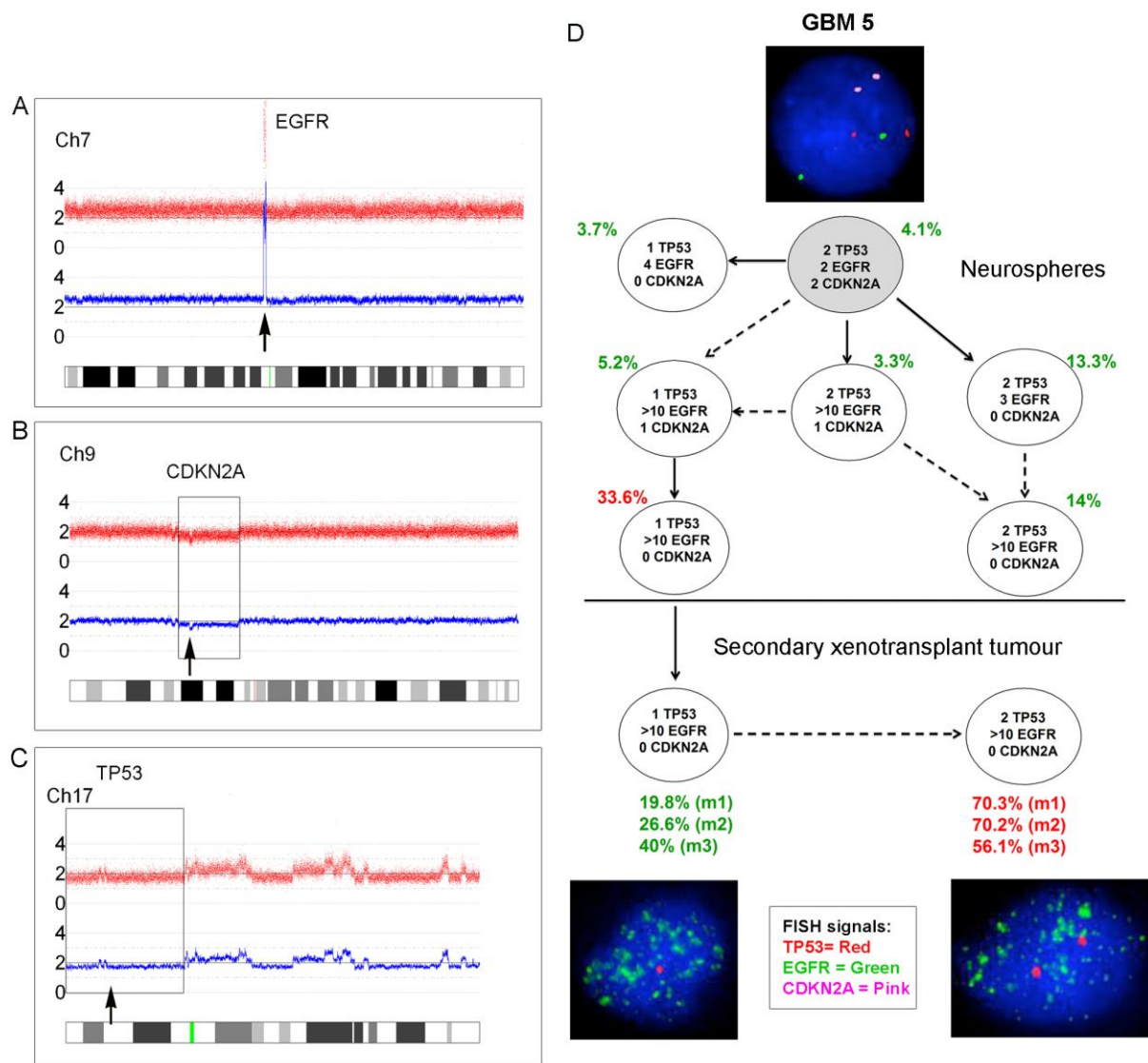


Figure S2, related to Figure 4. Sub-clonal genetic structure of neurosphere cells and tumour propagating cells derived from primary glioblastoma GBM 5 by multicolour FISH. (A-C) SNP 6 array profiles showing high level *EGFR* amplification (A), homozygous *CDKN2A* loss comprising a large deletion of one allele and a focal deletion of the second allele (B) and *TP53* loss due to a large 17p deletion (C): in each case the gene location is indicated by an arrow and the larger deletion by a box. (D) Clonal structure of neurosphere cells (top) and tumour cells after secondary transplant into NOD-SCID mice (bottom). FISH images are shown next to their respective genotype. Red type indicates the major clone; m1, m2, m3 = three replicate mice each injected with 1×10^6 neurosphere cells. The percentages of sub-clones with 1 or 2 or 3 copies of *TP53* determined by single cell analysis do not correspond to those determined by FISH (Figure 4). This is most likely due to the differences in probes used for the two methods. Solid arrows show probable derivation of sub-clones. Dashed arrows indicate possible alternative derivation of sub-clones. FISH images were captured at 100x magnification.

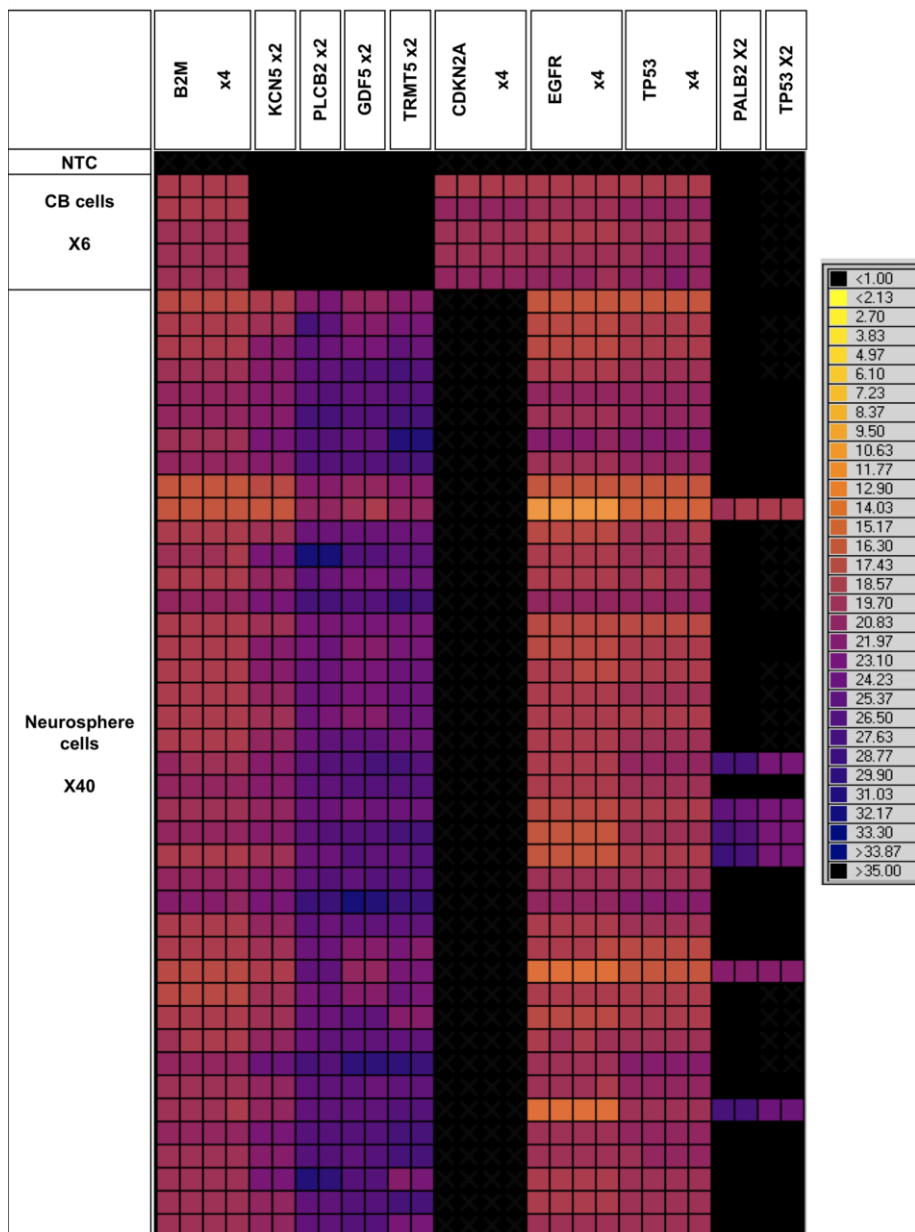


Figure S3, related to Figure 4. Heatmap depicting an example of raw Q-PCR data from the BioMark[®] HD. The rows represent single cells including six cord blood mononuclear cells and 40 neurosphere cells from GBM 5. The columns represent assays, each completed in quadruplicate including $\beta 2M$ (one of three assays). Mutations assessed were *KCN5*, *PLCB2*, *GDF5*, *TRMT5*, *PALB2* and *TP53*. Copy number alterations were *CDKN2A*, *EGFR* and *TP53*. The coloured boxes at the junction of a row and column indicate the raw C_T value (according to the key on the right) obtained for a Q-PCR reaction involving the indicated cell and assay. Assays targeting a mutation provide a definitive positive or negative result indicating the presence or absence respectively of an alteration (black = no mutation). The DNA copy number assays provide a raw C_T value which requires further analysis (standard $\Delta\Delta C_T$ method Applied Biosystems[®]) to attribute a DNA copy number to the target gene of interest for a single cell.

GBM 5

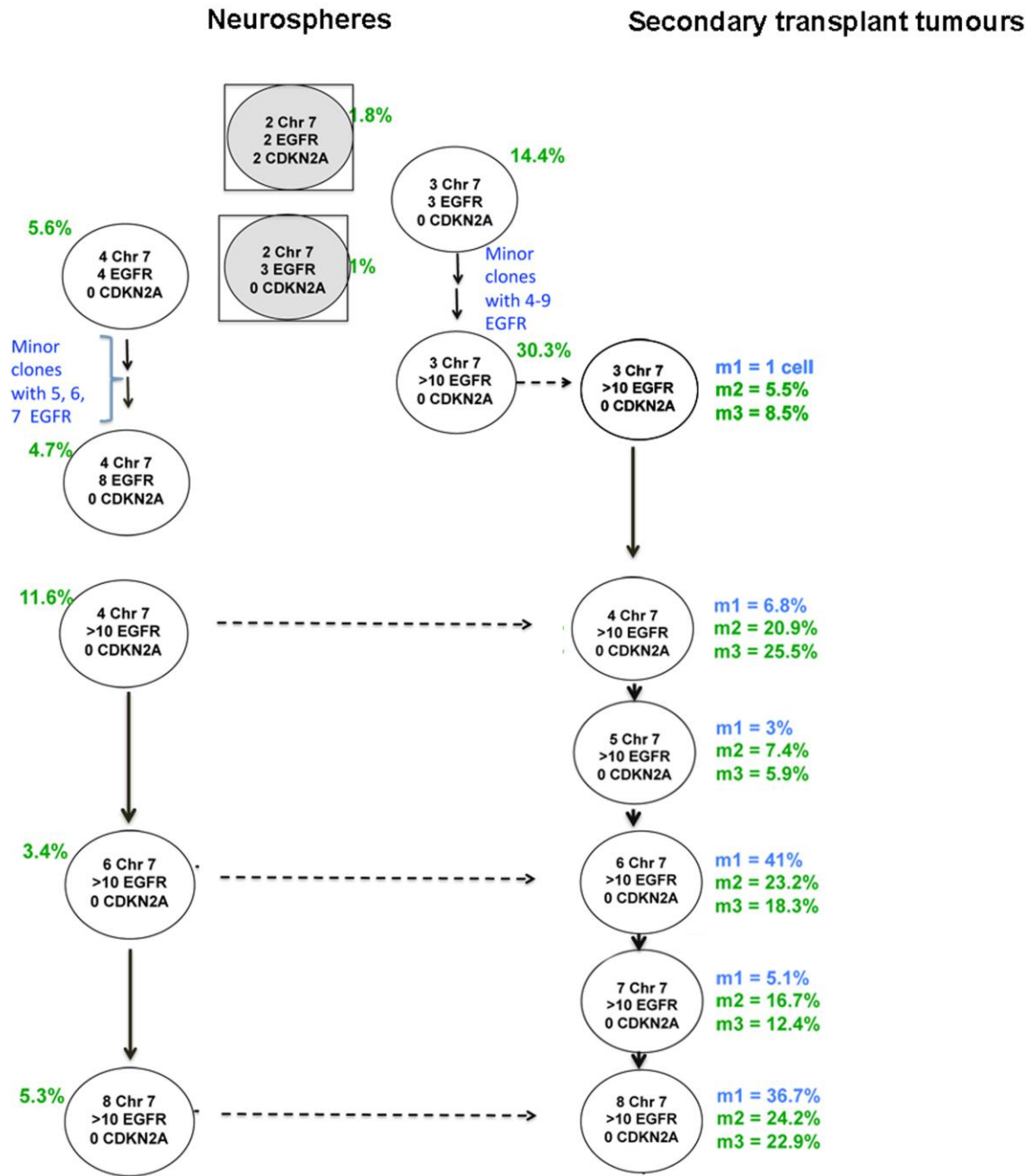


Figure S4, related to Figure 4. Sub-clonal genetic structure of GBM 5 neurospheres and secondary xenotransplant tumours using FISH probes for chromosome 7 copy number, *EGFR* amplification and *CDKN2A* loss. Stepwise clonal evolution with progressive gain of copies of chromosome 7 and *EGFR* amplification was observed in the neurospheres and in the tumours of secondary transplanted mice. *EGFR* amplification uncoupled from chromosome copy number at 3-4 copies of chromosome 7: thereafter *EGFR* amplification increased to very high levels, consistent with the formation of extrachromosomal double minutes (dmin). m1, m2, m3 = three replicate mice each injected with 1×10^6 neurosphere cells. Solid arrows show probable derivation of sub-clones. Dashed arrows indicate possible alternative derivation of sub-clones.

SUPPLEMENTAL TABLES

Table S1, related to Figure 1-4. Clinical information of patients.

Sample	Age at diagnosis (y)	Sex	Performance status	Treatment post-surgery
GBM 1	72	F	Not documented	SCRT
GBM 2	61	M	0	SCRT
GBM 3	61	M	0	SCRT
GBM 4	61	M	3	SCRT
GBM 5	52	M	0	ChemoRT
GBM 6	62	M	0	SCRT
GBM 7	63	F	1	ChemoRT
GBM 8	61	M	0	ChemoRT
GBM 9	63	F	1	SCRT
GBM 10	46	F	1	SCRT
GBM 11	64	F	0	ChemoRT
GBM 12	36	M	0	ChemoRT

We provide here clinical information of 12 GBM patients. The samples obtained from these patients were used in the study. These patients were prospectively recruited through the Neuro-oncology multidisciplinary team (MDT). All patients had suspected high-grade glioma on imaging review by a consultant neuroradiologist at the MDT. This study included 7 men and 5 women. Mean age was 58.5 years at diagnosis (range 36-72). WHO performance status before surgery was 0-3. Treatment post-surgery was either short course of radiotherapy (SCRT) or chemo- and radio-therapy followed by chemotherapy (ChemoRT).

Table S2, related to Figure 1-4. SNP 6 analysis of GBM tumours.

The table summarizing the SNP 6 data of all the GBM tumours is provided as Excel file.

Table S3, related to Figure 1-4. Details of tumours in mice and whole-exome sequencing results of GBM 5 tumour.

	DAYS TO TUMOUR FORMATION					
	Primary transplant			Secondary transplant		
	mouse1	mouse 2	mouse 3	mouse1	mouse 2	mouse 3
GBM 1	159	164	164	107	133	119
GBM 2	149	164	164	43	133	133
GBM 3	163	163	163	121	121	121
GBM 4	124	124	133	77	96	166
GBM 5	147	149	149	92	69	69
GBM 6	65	65	65	28	119	107
GBM 7	119	119	119	36	149	134
GBM 8	105	113	105	71	50	52
GBM 9	120	120	120	35	35	29
GBM 10	159	189	210	n.t.f.	n.t.f.	n.t.f.
GBM 11	138	145	145	93	54	93
GBM 12	218	n.t.f.	n.t.f.	n.t.f.	n.i.	n.i.

This table summarizes the data from the *in vivo* transplantation in immunosuppressed animals. Two out of twelve GBM failed to form tumour either after the first transplantation (GBM 12, two mice out of three) or after the second transplantation (GBM 10 and GBM 12). Although we observed a trend towards shorter time to tumour formation in the secondary transplant compared to the primary xenograft tumour, a statistically significant difference was observed only in GBM 1, 5, 8, 9, 11 ($p < 0.05$). "n.t.f."= no tumour formation, "n.i."= not injected.

Chr.	SNV position	CNV Ref->Seq	Gene	Estimated allele burden (%)
1	29379636	G -> T	<i>EPB41</i>	28
1	57258326	G -> C	<i>C1orf168</i>	20
2	27549589	G -> A	<i>GTF3C2</i>	19
2	180348097	G -> A	<i>ZNF385B</i>	18
2	233321642	G -> A	<i>ALPI</i>	19
3	110611201	C -> A	<i>RP11-553A10.1</i>	24
6	49494443	G -> A	<i>GLYATL3</i>	27
7	16566662	C -> T	<i>LRRC72</i>	30
7	135387597	T -> C	<i>SLC13A4</i>	25
10	5966417	C -> A	<i>FBXO18</i>	22
10	79782048	G -> A	<i>POLR3A</i>	33
10	90427142	G -> C	<i>LIPF</i>	30
10	124339169	G -> T	<i>DMBT1</i>	25
11	55587748	G -> A	<i>OR5D18</i>	19
11	56128024	T -> A	<i>OR8J1</i>	22
14	61446242	C -> T	<i>TRMT5</i>	28
14	63269168	T -> A	<i>KCNH5</i>	32
15	40591062	G -> A	<i>PLCB2</i>	24
15	79339105	G -> T	<i>RASGRF1</i>	20
15	79748827	G -> A	<i>KIAA1024</i>	17
16	23635396	A -> G	<i>PALB2</i>	22
16	81934351	C -> T	<i>PLCG2</i>	24
16	82033438	G -> A	<i>SDR42E1</i>	19
17	7578476	G -> A	<i>TP53</i>	20
17	61623193	C -> T	<i>KCNH6</i>	36
19	8577520	C -> T	<i>ZNF414</i>	17
19	35832277	C -> G	<i>CD22</i>	17
19	36211360	G -> A	<i>MLL4</i>	20
19	45899676	C -> T	<i>PPP1R13L</i>	22
20	34022533	C -> T	<i>GDF5</i>	15
8	105263344	C -> T	<i>RIMS2</i>	15
9	100074426	C -> T	<i>C9orf174</i>	17

Mutation targets selected for single cell interrogation are in bold.
An allele burden of 50% indicates either a heterozygous mutation in every cell or a homozygous mutation in 25% of cells.

Table S4, related to Figure 1-4. SNP Array 6.0 call rates and primer sequences for CE-SSCA and Sequencing Analysis of *TP53*.

File ID	Gender	Call Rate (%)	Het Rate	Hom Rate
GL1.CEL	female	97.14651	27.00407	70.14243
GBM1.CEL	female	96.41983	22.32763	74.09221
GL2.CEL	male	97.09704	26.2055	70.89154
GBM2.CEL	male	97.57921	24.14025	73.43897
GL3.CEL	male	98.08635	26.00916	72.07719
GBM3.CEL	male	93.94089	28.9871	64.95379
GL4.CEL	male	97.65738	25.63307	72.02431
GBM4.CEL	male	95.93447	22.80398	73.13049
GL5.CEL	male	97.94585	26.28663	71.65922
GBM5.CEL	male	97.1453	25.80709	71.3382
GL6.CEL	male	95.51979	27.43678	68.08301
GBM6.CEL	male	94.57115	26.1391	68.43205
GL7.CEL	female	98.22553	27.00704	71.21848
GBM7.CEL	female	96.77086	27.22076	69.5501
GL8.CEL	male	98.17408	26.10953	72.06455
GBM8.CEL	male	96.87013	23.56089	73.30924
GL9.CEL	female	96.93477	27.78198	69.15279
GBM9.CEL	female	98.233	26.70879	71.52422
GL10.CEL	female	98.03644	26.57181	71.46463
GBM10.CEL	female	97.43432	25.67594	71.75838
GL11.CEL	female	98.33898	26.81795	71.52103
GBM11.CEL	female	96.73733	25.23356	71.50377
GL12.CEL	male	98.17144	25.81842	72.35302
GBM12.CEL	male	97.95739	24.96729	72.9901

GL, Germ Line; GBM, Glioblastoma tumour
 Hom, homozygosity; Het, heterozygosity

Primer Name	Primer sequence (5' to 3') Labelled and unlabelled primers	Label
TP53 ex4	TP53 ex4 Fw GACCTGGTCCTCTGACTGCT TP53 ex4 Rv GCATTGAAGTCTCATGGAAG	FAM
TP53 ex5	TP53ex5 Fw TGTCGGCTGACTTTCAACTCT TP53ex5Rv GGCAACCAGCCCTGTCGT	FAM
TP53 ex6	TP53 ex6 Fw GCTGGGGCTGGAGAGACGA TP53 ex6 Rv CTGGAGGGCCACTGACAAC	VIC
TP53 ex7	TP53 ex7 Fw GGTCTCCCCAAGGCCCACTG TP53 ex7 Rv GGGGATGTGATGAGAGGTGGAT	NED
TP53 ex8	TP53 ex8 Fw GCCTCTTGCTTCTCTTTTCCTATC TP53 ex8 Rv GGGAGAGGAGCTGGTGTGTT	PET
TP53 ex9	TP53 ex9 Fw AGCAGGACAAGAAGCGGTGG TP53 ex9 Rv AACGGCATTGAGTGTAGACTG	FAM

CE-SSCA uses fluorescently labelled primers. The same primer sequences but with a universal tag instead of a fluorescent tag were used for PCR set up of individual exons prior to sequencing.

Universal tag:

Universal20-Fw	GTT GTA AAA CGA CGG CCA GT
Universal20-Rv	CAC AGG AAA CAG CTA TGA CC

TP53 CE-SSCA PCR cycling conditions

95 °C 7mins
 94 °C 30 secs }
 60 °C 30 secs } x40
 72 °C 30secs }
 72 °C 7 mins }
 25 °C ∞

BIGDYE FAST sequencing conditions

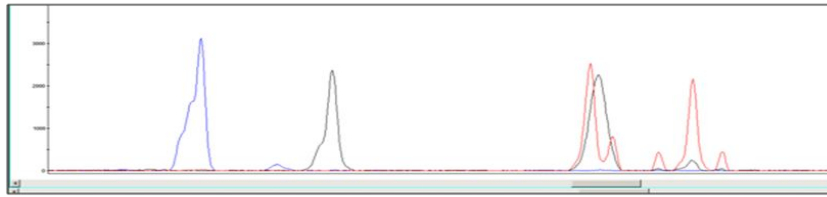
96 °C 1 min x1
 96 °C 10 secs }
 50 °C 5 secs x } 25
 60 °C 1 min 15 secs }
 4 °C ∞

Temperatures for CE-SSCA

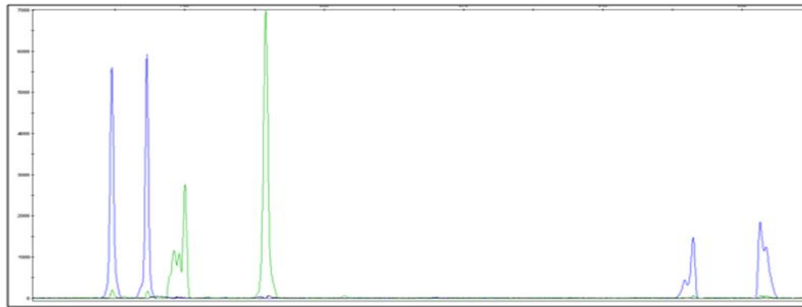
Temperature 1: 20 °C
 Temperature 2: 25 °C
 Temperature 3: 30 °C
 Temperature 4: 33 °C
 Temperature 5: 35 °C

TP53 schematic of SSCA results

TP53 SSCA analysis at 25°C



Mix 1: Ex 5 (FAM) Ex 07 (HEX) Ex 08 (PET)



Mix 2: Ex 4 (FAM) Ex 06 (HEX) Ex 04 (FAM)

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

DNA extraction

DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen) from the same GBMs used for cell line derivation after chopping and mixing of the tissue. Germline DNA was extracted with the same kit using blood or buffy coat. DNA quantification was performed using the picogreen assay for single-nucleotide polymorphism (SNP) analysis.

Copy number and LOH analysis

Samples were analyzed with Partek Genomics Suite 6.6 and CNAG 3.3.0.1 (beta) (<http://plaza.umin.ac.jp/genome/>) with the use of paired tumor (test) samples with the self-reference control (reference) samples to determine copy number and LOH caused by imbalance (Nannya et al., 2005). The position of regions of LOH and gain were identified with the University of California Santa Cruz Genome Browser (Hg18, March 2006 Assembly; <http://genome.ucsc.edu/cgi-bin/hgGateway>).

Genome mapping analysis

Genome-Wide Human SNP Array 6.0 (Affymetrix) mapping analysis was performed with 500 ng of tumour and germline DNA from each patient essentially as previously described. Briefly, genomic DNA was digested in parallel with restriction endonucleases NspI and StyI, ligated to an adaptor, and subjected to PCR amplification with adaptor-specific primers. The PCR products were digested with DNaseI and labeled with a biotinylated nucleotide analog. The labeled DNA fragments were hybridized to the microarray, stained by streptavidin-phycoerythrin conjugates, washed with the Affymetrix Fluidics Station 450, and then scanned with a GeneChip scanner 3000 7G. The CEL files are deposited with Gene Expression Omnibus, GEO Accession no. GSE45185. SNP genotypes were obtained with the use of the BRLMM algorithm in Affymetrix Genotyping Console 4.0 software (see Table S4 for the SNP Array 6.0 call rates).

Fluorescence *in situ* hybridization

In each case, 100 - 200 nuclei were scored for the presence of the relevant probe signals. Depending on the probe, amplified gene signals were scored manually up to 5-9 signals: high-level amplifications were recorded as > 10 signals. Cut-off levels for false loss of gene signals were established using normal peripheral blood control slides as previously described (Anderson et al., 2011). For *CDKN2A*, a threshold cut-off was set at 2% for loss of a single signal and 1% for loss of two signals. Clonal architectures were established by direct scrutiny of the degree of similarity of genotypes, with the assumption that acquisition of CNA is irreversible and genotypes change unidirectionally.

Detection of *TP53* Mutations by Capillary Electrophoresis Single-Strand Conformation Analysis

We used a capillary electrophoresis single-strand conformation analysis (CE-SSCA) method to detect mutations in exons 4-9 of the *TP53* gene. Any mutations identified were further characterised by Sanger sequencing.

We used a capillary electrophoresis single-strand conformation analysis (CE-SSCA) approach for identification of mutations in exons 4-9 of the *TP53* gene. CE-SSCA involves PCR amplification using fluorescently labelled primers followed by electrophoresis through a non-denaturing capillary across a range of temperatures to increase sensitivity. *TP53* exons 5, 7 and 8 are amplified in a PCR multiplex and each primer pair is labelled with a different fluorochrome (Table S4). The PCRs for exons 4, 6 and 9 are set up individually (using of two different fluorochromes) and then multiplexed prior to CE-SSCA electrophoresis on an ABI 3130xl Genetic Analyser.

The temperatures used for electrophoresis on the ABI 3130xl were 20, 25, 30, 33 and 35°C. The CE-SSCA data was analysed using Genemapper Software. The PCR product for each exon produces a specific pattern of peaks at each different temperature. The pattern for the wild type sample at each temperature was used as a reference for

analysis of patient samples. Any mobility shift identified by CE-SSCA was subsequently further tested by Sanger sequencing to elucidate the precise mutation.

For Sanger sequencing, a second PCR was performed for the specific exon (exon 5 for GBM 5 and exon 7 for GBM 8). The primer sequences used for this PCR are the same as the CE-SSCA but instead of a fluorescent label the primers are universally tagged. The PCR products were then cleaned-up using Ampure (supplied by Agencourt) and sequenced using BigDye Terminator v3.1. Ethanol, EDTA and Sodium Acetate precipitation was then performed prior to running the samples on the ABI 3730 capillary sequencer. Sequencing data was analysed using Mutation Surveyor software.

Whole-exome sequencing

Briefly, DNA was sheared by fragmentation (Covaris, Woburn, MA, USA) and purified using Agencourt AMPure XP beads (Beckman Coulter, Fullerton, CA, USA). The resulting fragments were analysed using an Agilent 2100 Bioanalyzer. Fragment ends were repaired and adaptors were ligated to the fragments. The library was purified using Agencourt AMPure XP beads and amplified by PCR before hybridisation with biotinylated RNA baits. Bound genomic DNA was purified with streptavidin coated magnetic Dynabeads (Invitrogen, Carlsbad, CA, USA) and re-amplified to include barcoding tags before pooling for sequencing on a paired-end, 100 cycle run on an Illumina HiSeq 2000 according to manufacturer's protocols. Somatic SNPs were identified using SomaticSniper (v1.0) (Larson et al., 2012) and were restricted to being called in regions known to be targeted by the exome capture kit. CNV analysis was performed with VarScan2 (Koboldt et al., 2012) using default parameters.

Single cell sorting

Single cell sorting was performed on a BDFACS Aria1-SORP instrument (BD®, Franklin Lakes, NJ, USA) equipped with an automated cell deposition unit using the following settings: 100micron nozzle, 1.4bar sheath pressure, 32.6KHz head drive and a flow rate that

gave 1-200 events per second. Viable cell thawing, single cell carboxyfluorescein diacetate N-succinimidyl ester (CFSE) staining of the neurosphere sample (according to manufacturer's instructions), cell sorting parameter explanations and the assessment of single cell sorting efficiencies was completed as previously described (Potter et al., 2013). Only fixed material surplus to FISH requirements was available from the xenograft transplant. This material was resuspended in PBS and labeled with propidium iodide prior to FACS. Five plates of single cells were collected from the GBM 5 neurosphere sample and two from the fixed xenograft material each composed of a no template control (NTC), 11 control cord blood cells and 84 target cells (GBM 5).

CNA assays were completed in quadruplicate and SNV assays were completed in duplicate. A heterozygous mutation was considered to be present if the Q-PCR C_T value from the mutant and wild-type sequence probes (FAM and VIC respectively) was <28 for a single cell. A homozygous mutation was considered to be present if there was no wild-type sequence signal. To ensure robust DNA CNA data from a system that can be influenced by assay efficiency and experimental variation we employed the $\Delta\Delta C_T$ method (Applied Biosystems[®]) to determine a copy number for each locus with modifications to incorporate data from three distinct assays targeting the control region (*B2M*) and the region of interest. Details of this approach can be found in (Potter et al., 2013).

Single cell multiplex targeted pre-amplification and Q-PCR

Labelled single cells were sorted into 2.5 μ l lysis buffer composed of 1 mg/ml proteinase K (Qiagen Ltd, Manchester, UK) and 0.5% Tween20 in HEPES buffered saline (Sigma-Aldrich[®], Gillingham, UK). Lysis was carried out for 50 min at 60°C followed by 10 min at 98°C. Specific (DNA) targeted amplification (STA) was then performed prior to Q-PCR. This multiplex STA reaction was composed of 5 μ l pre-amplification master mix (Life Technologies Ltd) and 2.5 μ l 1:40 primer mix (containing all primers for the gene targets of interest designed in house according to ABI guidelines). Denaturation was completed at 95°C for 15 min, followed by 24 cycles of amplification at 95°C for 15 sec and 60°C for 4 min. The STA

product was then diluted 1:6 using DNA suspension buffer (TekNova[®], Surrey, UK). Finally, 2.7 μ l of the single cell target amplified DNA was interrogated by Q-PCR for each DNA target of interest using the 96.96 dynamic microfluidic array and the BioMark™ HD as recommended by the manufacturer; thermal phase 70°C for 1800 sec, 25°C for 60°C sec followed by a hot start phase of 95°C for 60 sec. This was followed by 35 cycles of 96°C for 5 seconds and 60°C for 20 sec. CNA assays were completed in quadruplicate and SNV assays were completed in duplicate.

Single cell Q-PCR analysis

The BioMark™ HD generates a C_T value for each reaction. A heterozygous mutation was considered to be present if the signals from the mutant and wild-type sequence probes (FAM and VIC respectively) had a C_T value <28 in a single cell. A homozygous mutation was considered to be present if there was no wild-type sequence signal. To ensure robust DNA CNA data from a system that can be influenced by assay efficiency and experimental variation we employed the $\Delta\Delta C_T$ method (Applied Biosystems[®]) to determine a copy number for each locus with modifications to incorporate data from three distinct assays targeting the control region (*B2M*) and the region of interest. The $\Delta\Delta C_T$ value was calculated for every target gene assay using each of the three reference gene C_T values generating nine estimated DNA copy number results for a region of interest. A confidence metric was assigned to the estimated copy number inferring the confidence with which an estimated copy number could be deemed true (according to Applied Biosystems CopyCaller[®] Software v2). Details of this approach can be found in Potter et al., 2013. The weighted mean of the nine estimated DNA copy numbers (for a region of interest) was used as the final DNA copy number taking into consideration the confidence metric attributed to each. This reduced the contribution of less reliable estimated DNA copy numbers to the final DNA copy number. Estimated copy number results were not considered if the confidence value was less than 50% or the estimated copy number was greater than four (with only quadruplicates per

assay the results are not robust enough to accurately detect DNA copy numbers greater than four. At least two of the nine estimated copy numbers must have a confidence value above 50% to calculate the final copy number for a region of interest.

SUPPLEMENTAL REFERENCES

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research* 22, 568-576.

Larson, D.E., Harris, C.C., Chen, K., Koboldt, D.C., Abbott, T.E., Dooling, D.J., Ley, T.J., Mardis, E.R., Wilson, R.K., and Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 28, 311-317.

Nannya, Y., Sanada, M., Nakazaki, K., Hosoya, N., Wang, L., Hangaishi, A., Kurokawa, M., Chiba, S., Bailey, D.K., Kennedy, G.C., *et al.* (2005). A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer research* 65, 6071-6079.