SUPPLEMENTARY METHODS

Differential gene expression analysis of paediatric GBM

Expression data using the Affymetrix U133Plus2.0 platform from the Schwartzentruber\(^2\) and Paugh\(^3\) studies of paediatric GBM specimens were retrieved from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/). Data were re-normalized from original .CEL files using \texttt{rma} in R 2.15.1 (www.r-project.org/). \(H3F3A\) mutation status determined by Sanger sequencing (below) where not otherwise reported. Differential expression signatures were generated in GenePattern\(^28\) using a signal-to-noise metric, with a cut-off of 0.85 for assigning genes to a mutation specific signature. Gene set enrichment analysis\(^29\) was used to probe for coordinated upregulation of mutation signatures in complementary datasets. A Pre-ranked analysis was used for probing enrichment of paediatric GBM signatures against the ChIP-Seq data. A set of core enrichment genes was identified as overlapping between gene expression and ChIP-Seq data, defined as those genes ranked at or higher than the gene at which the enrichment score peaks. Consensus k-means clustering of paediatric GBM expression data was performed in GenePattern using a kmax of 10.

Mapping gene expression signatures to brain development

Expression data from Kang \emph{et al.}\(^9\) were retrieved from GEO and re-normalized from original .CEL files using \texttt{rma} in R 2.15.1. This data includes transcriptomes of 16 regions comprising the cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, hippocampus and 11 areas of the neocortex from 1,340 tissue samples collected from 57 developing and adult post-mortem brains. Paediatric GBM expression signatures associated with individual mutations were mapped to specific developmental stages and anatomic locations by taking the median expression of all signatures genes at each spatio-temporal point and visualized in R.
**Immunohistochemistry**

Tissue microarrays of 46 paediatric and young adult (median age = 14 years, range 6 months – 38.5 years) GBM samples were screened for H3F3A mutation status by direct sequencing (below) as well as protein expression of DLX6, SOX2, MYCN by immunohistochemistry. The rabbit EnVision™+ HRP System, (DAKO, Carpinteria, CA, USA) was used for SOX2 (2683-1, Epitomics, Burlingame, CA, USA) and DLX6 (NBP1-85929, Novus Biologicals, Cambridge, UK). Antigen retrieval for SOX2 was performed by pressure cooking in 10mM citrate buffer pH6 for 3 min, whilst boiling the slides for 20 min in 10mM citrate buffer pH6 was used for DLX6. Each section was blocked with DAKO Protein block (X0909) for 10 min at room temperature and incubated with primary antibody overnight at 4°C (SOX2 1:200, DLX6 1:50). The Universal R.T.U. Vectastain Elite ABC Kit (Vector Laboratories, Burlingame, CA, USA) was used for MYCN (sc-791, Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA). Antigen retrieval was performed by boiling the slides for 20 min in DAKO Target Retrieval Solution (S1699). After blocking for one hour, each section was incubated with primary antibody overnight at 4°C (MYCN 1:1000). Staining was completed with a 5 minute incubation with DAKO 3,3’-diaminobenzidine (DAB)+ substrate-chromogen and counterstained with Mayer’s Haematoxylin (Sigma-Aldrich, Poole, UK).

**H3F3A mutation screening**

DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) samples using the QIAamp DNA Micro Kit (56304) (Qiagen, Crawley, UK), after overnight incubation of 1M NASCN. Primers for H3F3A exon 1 were designed to generate PCR fragments of 181bp, suitable for FFPE specimens, using the >gi|224589800|ref|NC_000001.10| Homo sapiens chromosome 1, GRCh37.p5 Primary Assembly and were obtained from Life Technologies (Paisley, UK) as follows: Forward-
Each 25μl PCR reaction contained 2.5μl 10x buffer, 1.5μl 50mM MgCl₂, 2.5μl 2mM dNTPs, 0.5μl 10μM H3F3A forward primer, 0.5μl 10μM H3F3A reverse primer, 0.1μl Platinum Taq (Life Technologies), 7μl ddH₂O and 10μl DNA (10ng/μl). The PCR protocol included 35 cycles with denaturing at 95°C for 30 sec, annealing at 55°C for 40 sec, then extending at 72°C for 50 sec. The PCR products were visualised under UV on a 2% agarose gel stained with ethidium bromide and sent for sequencing. Sequences were analysed using Mutation Surveyor (SoftGenetics, Pennsylvania, USA) and manually with 4Peaks (Mekentosj, Aalsmeer, The Netherlands).

**Cell lines**

Cells were grown as monolayers in DMEM/F12 Ham’s medium + 10% FCS in 5% CO₂. KNS42 cells were found by direct Sanger sequencing to harbour the H3F3A G34V mutation (c.104G>T, p.Gly34Val), whilst all other cell lines were H3F3A and HIST1H3B wild-type.

**ChIP-Seq analysis**

ChIP was carried out using antibodies against H3K36me3 and RNA polymerase II using the HistonePath™ and TranscriptionPath™ assays by ActiveMotif. Whole genome sequencing was carried out using an Illumina HiSeq2000 instrument with a fold coverage of >30 fold. ChIP-Seq data was processed in R using Bioconductor 2.10 (www.bioconductor.org) libraries RSamtools, chipseq, GenomicRanges, htSeqTools and DESeq. ChIP-Seq reads were aligned to the genome using bwa with default parameters. Binary alignments were analysed for differential binding at both peak (region) level and gene level. Peaks were identified by slicing ChIP-Seq coverage at a minimum level of 5 reads. Peaks with a minimum width of 1kb were selected for differential analysis. For gene level coverage overlaps of each alignment
with 28729 Ensembl genes (assembly 64) were summarised as total reads, maximum and median coverage. Coverage statistics were corrected for input copy number in each case by removing the fraction of each peak defined by the median cbs smoothed log ratio covering each genomic region derived from Affymetrix 500K mapping arrays\textsuperscript{7} using the aroma Affymetrix CRMA method. Differential binding of corrected total read counts in each case was compared using the DESeq package. Differential genes were identified as those with an un-adjusted DESeq p value of less than 0.05 and an overall fold change of greater than 2 and a contiguous median coverage of greater than 2. Validation of selected active regions was carried out by ChIP-qPCR.

\textit{siRNA screening}

siRNA screening was carried out on a library of 714 human kinases using Dhharmacon SMARTpools\textsuperscript{TM} (Dharmacon, Lafayette, CO, USA) with cells plated in 96 well plates and transfected 24 hours later with siRNA using Lipofectamine RNAiMax\textsuperscript{TM} (Invitrogen) as per manufacturer's instructions. Each siRNA plate was supplemented with 10 wells of siControl. Twenty four hours following transfection, cells were trypsinised and divided into three identical replica plates. Media was replenished after 48 hours and 96 hours, and cell viability was assessed after seven days using CellTiter-Glo\textsuperscript{TM} Luminescent Cell Viability Assay (Promega) as per manufacturer's instructions. Data from each cell line was processed as follows: the luminescence reading for each well on a plate was log2 transformed and expressed relative to the median luminescence value of all wells on the same plate (plate centering). This data was then normalised according to the median of the entire screen data, using the median absolute deviation (MAD) to estimate the true variation within each screen. This normalisation represented the effect of each SMARTpool in each cell line as a Z score and allowed the effects of each
SMARTpool on viability to be compared across the cell line panel. A Z score $\leq -3$ was taken as the significance threshold for reduced cell viability, representing three MADs from the median and approximating to three standard deviations. Only those siRNAs passing this threshold in KNS42 cells were further considered. Synthetic lethality hits were proposed where the difference in Z scores between KNS42 cells and each H3F3A wild-type cell line were $\geq 3$, and ranked according to the Z score in KNS42.

**siRNA validation and drug assays**

Individual ON-TARGETplus™ oligonucleotides for validation were obtained from Dharmacon for AURKA (Set LQ-003545-01. Oligo#1:J-003545-26; Oligo#2:J-003545-27; Oligo#3:J-003545-28; Oligo#4:J-003545-29). Cells were reverse transfected with 20nM siRNA using Lipofectamine RNAiMax™ (Invitrogen) as per manufacturer's instructions in 96 well plates and assayed for up to 96 hours post-transfection. Data for each oligonucleotide was normalized to a negative control non-targeting siRNA (AllStars Negative Control siRNA, Qiagen). The AURKA-selective small molecule inhibitor VX-689 (MK-5108) was obtained from Selleckchem (Houston, TX) and assayed for up to five days at concentrations of 0.1, 0.5 and 2.5µM. Effects on cell viability assessed by CellTiter-Glo™.

**Construction of isogenic H3F3A G34V mutant cells**

Gene specific primers were used to clone H3F3A wild type and G34V alleles from cDNA reverse-transcribed from the KNS42 pGBM cell line (H3F3A_For: CACCTCCAGCCGAAGGAGAAGG; H3F3A_Rev_mut_stop: TCCTCCTGAAGCAGGAGGAGAAGG). Additional non-homologous 5’ sequence within the forward primer (CACC) is used to allow directional cloning into the pENTR®D-TOPO® vector (Invitrogen, K2420-20), and mutation of the stop codon was engineered into the reverse primer sequence to allow read-through of the open reading frame into a C-terminal V5 tag within the destination expression vector. The
destination lentiviral expression vector was produced using MultiSite Gateway LR recombination of the pENTR/D-TOPO® vector (Invitrogen, K2400-20) containing the \textit{H3F3A} allele, the pENTR 5'/EF1\textit{\alpha}P construct and the pLenti6.4/R4R2/V5-DEST vector. Expression vector construct components are contained in ViraPower™ HiPerform™ Promoterless Gateway® Expression System (Invitrogen, A11146) and the associated lentiviral packaging constructs are from the ViraPower™ Bsd Lentiviral Support Kit (Invitrogen, K4970-00).

Expression plasmid was mixed with associated lentiviral packaging plasmids and transfected into HEK293T cells are per the manufacturer's instructions. Virus containing supernatants were collected 48 hours later, spun at 3000 RPM for 15 minutes at 4°C, then the supernatant was filtered through a Millex-HV 0.45μM filter. Alliquots of virus were stored at -80°C until ready for use. Lentiviral titer was determined to be 3.5 x10^6 – 4.0 x 10^6 viral particles/mL using the QuickTiter Lentivirus Titer Kit (Cell Biolabs, Inc. VPK-107). Transductions were performed as per the manufacturer's instructions. Briefly, cells were seeded in 6 well tissue culture plates at 2x10^5 cells/well (NHA cells) or 1x10^5 cells/well (SVG cells). The next day, 100μL of supernatant containing virus was added to cells in the presence of 4μg/mL polybrene (Sigma 107689). Twenty-four hours later virus containing media was replaced with fresh media and selection with blastocidin (Invitrogen, R21001) was initiated twenty-four hours after removal of the virus. SVG cells were selected with 4μg/mL blastocidin and NHA cells were selected with 6μg/mL blastocidin.

\textit{Quantitative RT-qPCR}

Reverse-transcription followed by quantitative PCR was performed using the EXPRESS One-Step Superscript® qRT-PCR Universal (Invitrogen, 11781200) on the Applied Biosystems 7500HT. Reactions contained 0.5μg RNA and individual
measurements of Human *MYCN* (Applied Biosystems, Taqman Gene Expression Assay 1160739) and Human *HPRT* endogenous control locus (Pre-Developed Taqman Assay Reagent, Applied Biosystems, 4326321E) were measured in triplicate.