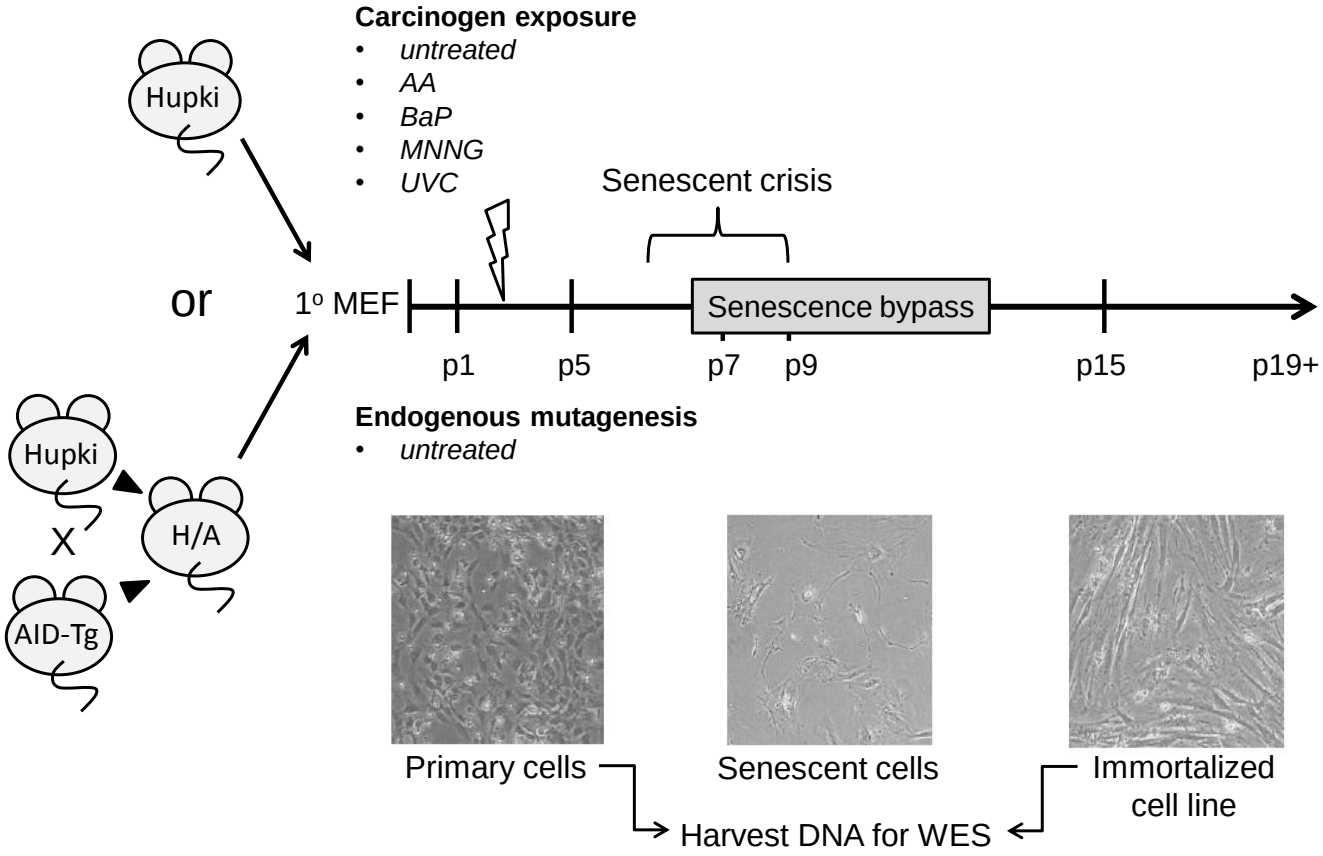
Supplementary Information

TITLE
Modelling mutational landscapes of human cancers *in vitro*

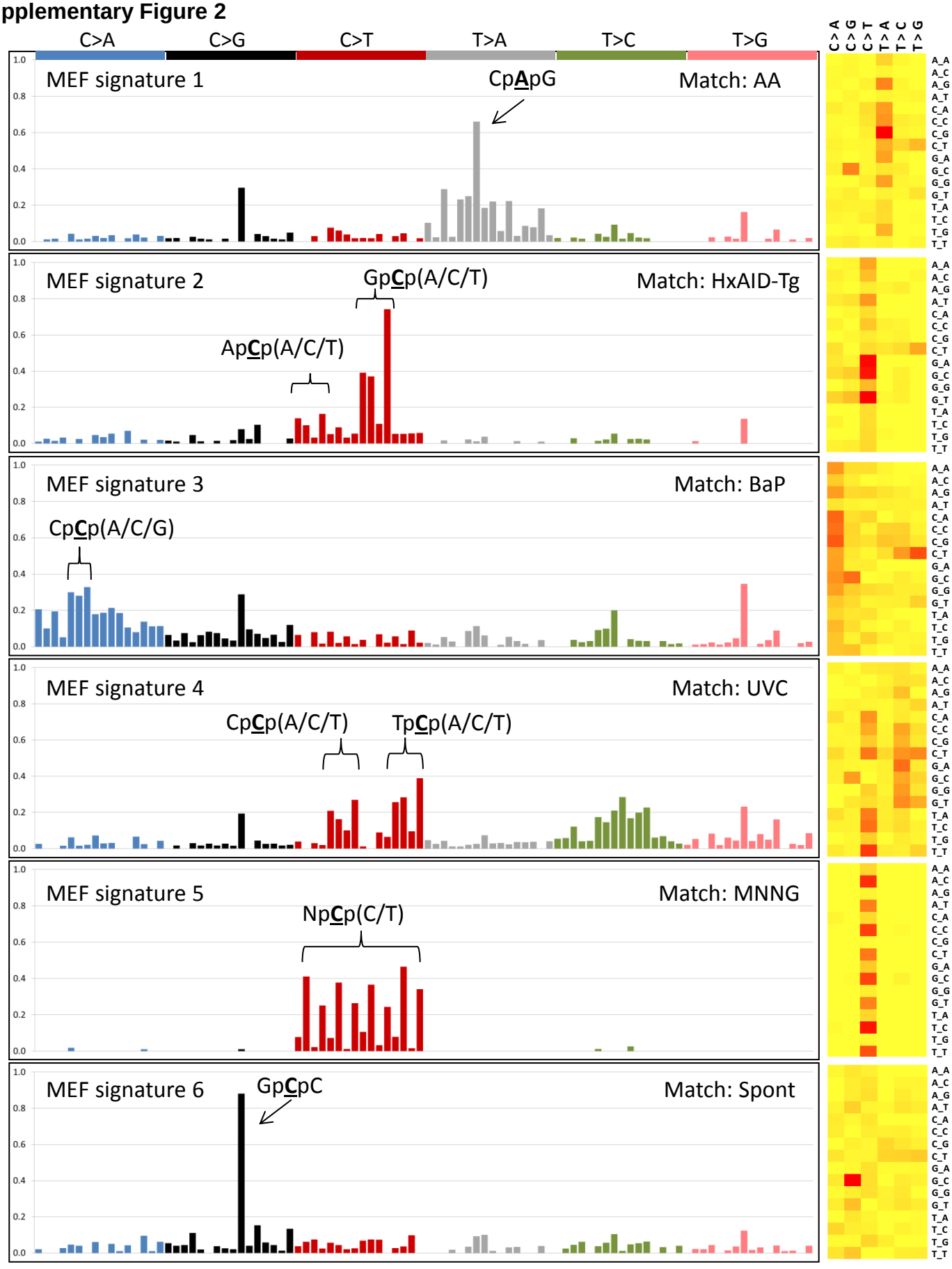
AUTHORS
Magali Olivier, Annette Weninger, Maude Ardin, Hana Huskova, Xavier Castells, Maxime P. Vallée, James McKay, Tatiana Nedelko, Karl-Rudolf Muehlbauer, Hiroyuki Marusawa, John Alexander, Lee Hazelwood, Graham Byrnes, Monica Hollstein and Jiri Zavadil

Supplementary Figure 1



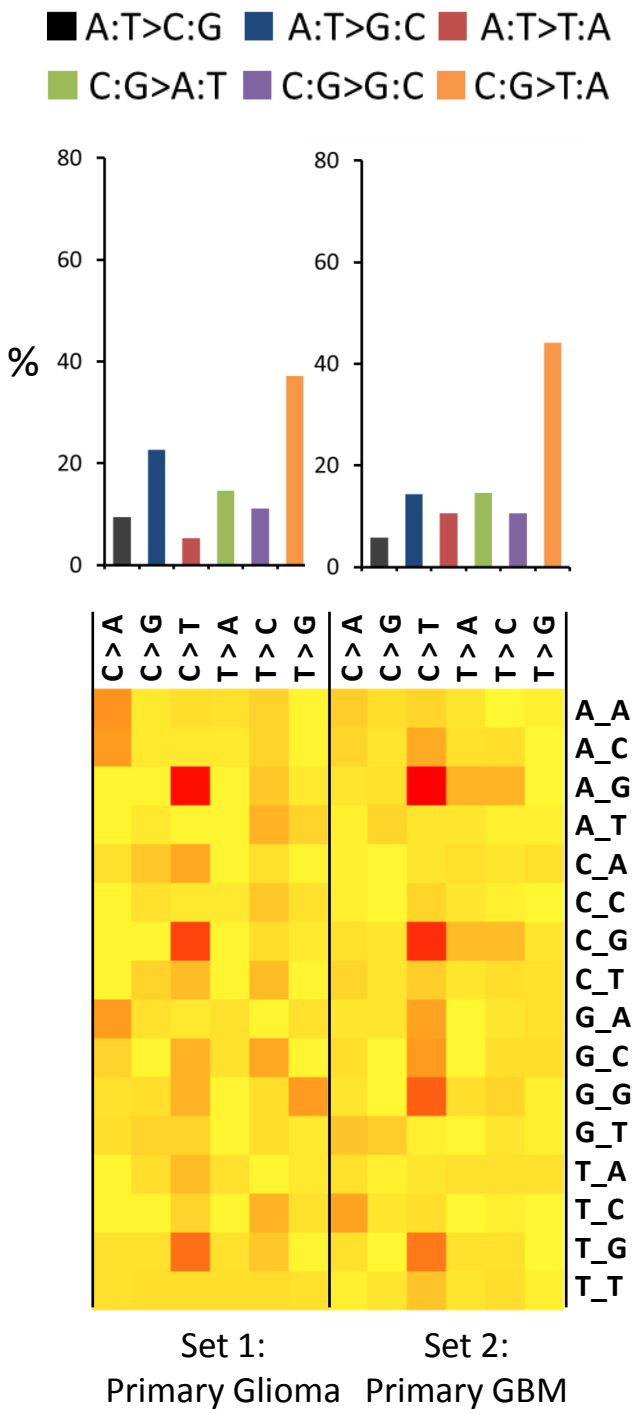
Supplementary Figure 1. Schematic overview of the modified 3T3 protocol for immortalisation of MEFs used to obtain cell lines for WES. Twenty-four or six cultures per embryo were used for experiments within each condition. p1 – p19+, arbitrarily selected cell culture passage numbers to facilitate orientation. The carcinogen exposure experiments included untreated controls and no substance treatment was applied in the assays for endogenous mutagenesis. Abbreviations used in the figure are as in the main text with the exception of H/A standing for the HxAID-Tg cross.

Supplementary Figure 2



Supplementary Figure 2: Signatures identified by non-negative matrix factorisation in 14 MEF cell lines and their match with the 6 experimental conditions (MEFs treated with AA, BaP, MNNG, UVC, endogenous mutagenesis sytem HxAID-Tg and spontaneously immortalised cell lines). Y-axis: frequency distribution. The heat map depiction of these signatures (right) allows a comparison with **Fig. 1** data.

Supplementary Figure 3



Supplementary Figure 3. Mutation patterns and single base substitution sequence context in primary gliomas from patients not exposed to temozolomide. Data were extracted from Set1: Johnson BE, et al. Science. 2014 Jan 10;343(6167):189-93. PubMed PMID: 24336570; Set2. Yost SE, et al. PLoS One. 2013;8(2):e56185. PubMed PMID: 23441165; PubMed Central PMCID: PMC3575368

Supplementary Table 1: List of samples, treatment conditions and the *TP53* mutation status.

Cell line ID	Exposure type	Exposure dose	Exposure duration	TP53 mutation	Zygosity	Reference
AA_1	AA	50 uM	4 days	c.745A>T (p.R249W)	homozygous	Nedelko et al., Int J Cancer (2009)
AA_2	AA	50 uM	4 days	c.391A>T (p.N131Y); c.871A>T (p.K291*)	heterozygous	Nedelko et al., Int J Cancer (2009)
MNNG_1	MNNG	20 uM	2 hrs	c.296C>T (p.S99F)	homozygous	Nedelko et al., Int J Cancer (2009)
MNNG_2	MNNG	20 uM	2 hrs	c.734G>A (p.G245D)	homozygous	Nedelko et al., Int J Cancer (2009)
BaP_1	BaP	1 uM	6 days	c.845G>C (p.R282P)	heterozygous	Reinbold et al., Oncogene (2008)
BaP_2	BaP	1 uM	6 days	c.423C>G (p.C141W)		Reinbold et al., Oncogene (2008)
UVC_1	UVC	20 J/m2	n.a.	c.743G>A (p.R248Q); c.747G>A (p.R249R); c.749C>A (p.P250H)	homozygous	Liu et al., PNAS (2004)
UVC_2	UVC	20 J/m2	n.a.	c.535C>T (p.H179Y)	homozygous	Liu et al., PNAS (2004)
HxAID-Tg_1	none	n.a.	n.a.	n.s.		unpublished
HxAID-Tg_2	none	n.a.	n.a.	n.s.		unpublished
Spont_1	none	n.a.	n.a.	n.s.		unpublished
Spont_2	none	n.a.	n.a.	n.s.		unpublished
Spont_3	none	n.a.	n.a.	c.314G>C (p.G105A)	homozygous	Whibley et al., J Biol Chem (2010)
Spont_4	none	n.a.	n.a.	c.843C>G (p.D281E)	homozygous	Whibley et al., J Biol Chem (2010)
Prim_1	none (primary MEF)	n.a.	n.a.	none		unpublished
Prim_2	none (primary MEF)	n.a.	n.a.	none		unpublished

n.a. = not applicable

n.s. = not sequenced

Supplementary Table 2: GO Biological Process of the C>G mutated genes in Spont cell lines

a

GO ID	GO Term	Gene Count	Fold Enrichment	% Enrichment	Fisher Exact p-value	Genes affected by non-synonymous C:G>G:C mutations
GO:0042981	**regulation of apoptosis**	**11**	**1.94**	**2.08**	**0.0260**	***IRAK3, CASP9, TRP53BP2, ADAMTSL4, GSK3B, CX3CR1, MAEL, NR4A1, BIRC6, COL2A1, SRC***
GO:0043067	**regulation of programmed cell death**	**11**	**1.92**	**2.08**	**0.0290**	***IRAK3, CASP9, TRP53BP2, ADAMTSL4, GSK3B, CX3CR1, MAEL, NR4A1, BIRC6, COL2A1, SRC***
GO:0007155	cell adhesion	11	1.92	2.08	0.0290	*HAPLN2, SVEP1, COL6A2, CDHR5, COL2A1, NLGN3, SCARB2, THBS1, SRC, EMILIN1, SCARF2*
GO:0010941	**regulation of cell death**	**11**	**1.91**	**2.08**	**0.0300**	***IRAK3, CASP9, TRP53BP2, ADAMTSL4, GSK3B, CX3CR1, MAEL, NR4A1, BIRC6, COL2A1, SRC***
GO:0007267	cell-cell signaling	10	3.37	1.89	0.0008	*FGFR4, RAPSN, P2RX3, P2RX2, GRIK5, SLC22A3, NLGN3, CACNA1C, SNAP25, FGF3*
GO:0007268	synaptic transmission	8	4.39	1.52	0.0005	*RAPSN, P2RX3, P2RX2, GRIK5, SLC22A3, NLGN3, CACNA1C, SNAP25*
GO:0019725	cellular homeostasis	8	2.28	1.52	0.0250	*CCR10, PTH1R, GRIK5, CACNA1G, GNPAT, NLGN3, CACNA1C, QSOX1*
GO:0043062	extracellular structure organization	7	4.59	1.33	0.0009	*ADAMTSL4, P2RX2, MYH11, GNPAT, COL2A1, NLGN3, EMILIN1*
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	6	2.54	1.14	0.0310	*SATB2, MAEL, SOX6, ZFPM1, TCF25, KDM6B*
GO:0032940	secretion by cell	5	2.63	0.95	0.0420	*EXOC7, CACNA1C, SYTL1, SNAP25, STXBP5L*

b

Refgene Symbol	Chr	Start	Ref	Alt	mRNA_refSeqID	nt_change	AA_change	Mapping	SBS mutation type	cytoBand	Genome Strand	Cell line
Adamtsl4	chr3	95487668	C	G	NM_144899	c.G911C	p.G304A	exonic	nonsynonymous	3qF2.1	-	Spont_2
Birc6	chr17	74965213	G	C	NM_007566	c.G1126C	p.A376P	exonic	nonsynonymous	17qE2	+	Spont_4
Casp9	chr4	141368196	G	C	NM_015733	c.G1237C	p.A413P	exonic	nonsynonymous	4qE1	+	Spont_2
Col2a1	chr15	97808836	G	C	NM_001113515	c.C3326G	p.S1109C	exonic	nonsynonymous	15qF1	-	Spont_4
Cx3cr1	chr9	119960659	C	G	NM_009987	c.G794C	p.S265T	exonic	nonsynonymous	9qF4	-	Spont_4
Gsk3b	chr16	38194009	G	C	NM_019827	c.G742C	p.A248P	exonic	nonsynonymous	16qB3	+	Spont_3
Irak3	chr10	119602212	G	C	NM_028679	c.C863G	p.A288G	exonic	nonsynonymous	10qD2	-	Spont_1
Mael	chr1	168131788	C	G	NM_175296	c.G1196C	p.S399T	exonic	nonsynonymous	1qH2.3	-	Spont_4
Nr4a1	chr15	101101224	G	C	NM_010444	c.G709C	p.A237P	exonic	nonsynonymous	15qF2	+	Spont_3
Src	chr2	157294960	G	C	NM_001025395	c.G1213C	p.A405P	exonic	nonsynonymous	2qH1	+	Spont_2
Trp53bp2	chr1	184378624	G	C	NM_173378	c.G2038C	p.E680Q	exonic	nonsynonymous	1qH5	+	Spont_2

Supplementary Table 2: (**a**) Gene Ontology analysis of the genes affected by non-synonymous C:G>G:C mutations in the spontaneously immortalised cell lines, using the NIH DAVID GO tool. (**b)** List of mutated genes and their acquired alterations that map to the highest enriched categories of regulation of apoptosis (GO:0042981) and regulation of programmed cell death (GO:0043067).

Supplementary Table 3: Summary metrics of WES data.

Sample	Total Number of Reads*	% Mapped reads, both in pair*	% Duplicates reads**	% Reads with quality > 20***	Mapped reads on target*	% Mapped reads on target*	% GC *	Mean Coverage*	% covered 1X*	% covered 14X*	% covered 30X*	Total number SBSs#	Pearson coefficient##
AA_1	55,116,881	95.78	26.83	90.58	44,082,924	79.98	54.81	71.86	99.026	83.290	63.215	644	0.6664
AA_2	33,201,370	99.34	5.36	95.32	26,500,867	79.82	51.38	42.33	99.667	85.970	56.826	588	0.6954
BaP_1	77,269,160	98.73	6.92	88.23	47,454,568	61.41	51.89	76.27	99.809	94.269	79.087	1206	0.535
BaP_2	48,600,495	99.34	6.08	81.64	37,907,317	78	51.4	60.52	99.720	91.264	70.903	671	0.4393
HxAID-Tg_1	49,447,050	98.54	7.63	97.21	38,069,266	76.99	49.5	62.39	99.742	89.523	65.681	293	0.5853
HxAID-Tg_2	48,452,293	98.57	5.87	97.39	37,663,039	77.73	49.51	62.11	99.734	88.697	64.152	550	0.6105
MNNG_1	54,535,240	96.69	30.69	79.51	45,342,126	83.14	54.36	73.63	99.030	85.196	66.084	1196	0.5438
MNNG_2	54,527,486	97.29	32.54	80.69	41,517,397	76.14	53.94	67.09	98.941	84.058	63.509	1087	0.5926
Spont_1	54,140,927	98.33	7.04	97.41	42,311,420	78.15	49.54	69.59	99.651	91.558	70.805	207	0.4396
Spont_2	53,784,459	98.42	6.33	97.25	41,715,179	77.56	49.08	68.72	99.631	90.406	68.229	291	0.4841
Spont_3	50,771,299	97.25	31.36	84.95	42,049,679	82.82	54.32	68.01	99.126	83.987	63.773	109	0.53
Spont_4	55,843,452	99.32	5.55	78.28	44,513,764	79.71	52	71.44	99.775	93.123	76.303	321	0.3561
UVC_1	9,481,442	97.72	5.73	96.32	5,239,673	55.26	50.83	8.3	95.924	16.626	1.888	346	0.5069
UVC_2	38,085,395	99.48	6.39	86.97	30,255,889	79.44	51.03	48.03	99.623	87.900	62.282	937	0.5232
Prim_1	55,188,132	98.18	5.88	98.18	41,878,241	75.88	49.66	68.54	99.72	92.21	71.64	NA	NA
Prim_2	46,200,087	97.15	30.58	88.2	38,651,510	83.66	54.22	62.66	99.10	83.22	62.03	NA	NA

* Metrics obtained with QualiMap using BAM files

** Metrics obtained with BamUtil from duplicates marked by Picard in BAM file

*** Metrics obtained with BamUtil using BAM files

\# Total number of SNV considered after variant calling by MuTect and filtering as described in Methods

Pearson correlation coefficient between the number of SNV per chromosome and chromosome size, calculated with R

Supplementary Datasets Captions

Supplementary Dataset 1 (xls file): List of all annotated single base substitutions found in oncogenes, tumour suppressor genes and regulators of the epigenome.

Supplementary Dataset 2 (xls file): Summary of non-synonymous and truncating single base substitutions found in oncogenes, tumour suppressor genes and regulators of the epigenome. Conditions (exposures or AID transgenic status) and corresponding cell lines are shown in column headings (abbreviations are used as in main text). The numbers of mutations in each cell line are shown for individual genes and the fields are colour-coded to indicate condition-specific predominant alteration (yellow) and non-specific alteration (blue), or presence of both (yellow/blue in diagonally split fields). Asterisk denotes that a gene is listed in at least two functional categories.

Supplementary Dataset 3 (xls file): Lists of all annotated single base substitutions identified in MEF cell lines and in used human datasets.