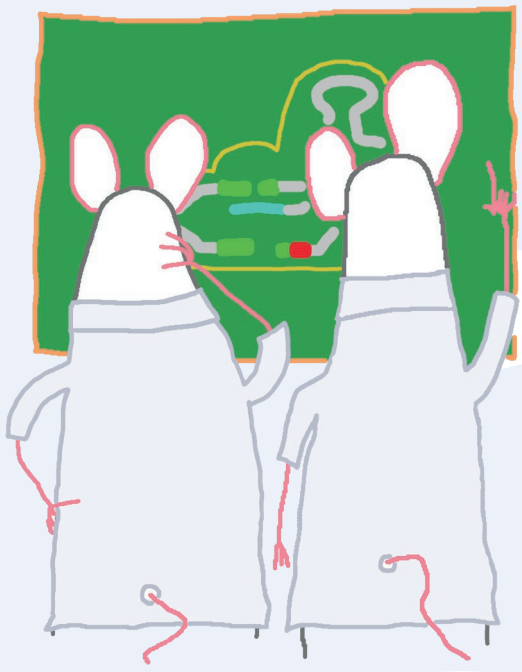
Blueprints of disease: precision platforms for modelling breast cancer

Lutz, C.



CHAPTER 5

In situ CRISPR-Cas9 base editing for the development of genetically engineered mouse models of breast cancer

Stefano Annunziato*, Catrin Lutz*, Linda Henneman, Jinhyuk Bhin, Kim Wong, Bjørn Siteur, Bas van Gerwen, Renske de Korte-Grimmerink, Maria Paz Zafra, Emma M Schatoff, Anne Paulien Drenth, Eline van der Burg, Timo Eijkman, Siddhartha Mukherjee, Katharina Boroviak, Lodewyk FA Wessels, Marieke van de Ven, Ivo J Huijbers, David J Adams, Luke E Dow & Jos Jonkers

* shared first authors

EMBO Journal. 2020 March 2;39(5):e102169

Abstract

Genetically engineered mouse models (GEMMs) of cancer have proven to be of great value for basic and translational research. Although CRISPR-based gene disruption offers a fast-track approach for perturbing gene function and circumvents certain limitations of standard GEMM development, it does not provide a flexible platform for recapitulating clinically relevant missense mutations *in vivo*. To this end, we generated knock-in mice with Cre-conditional expression of a cytidine base editor and tested their utility for precise somatic engineering of missense mutations in key cancer drivers. Upon intraductal delivery of sgRNA-encoding vectors, we could install point mutations with high efficiency in one or multiple endogenous genes *in situ* and assess the effect of defined allelic variants on mammary tumourigenesis. While the system also produces bystander insertions and deletions that can stochastically be selected for when targeting a tumour suppressor gene, we could effectively recapitulate oncogenic nonsense mutations. We successfully applied this system in a model of triple-negative breast cancer, providing the proof of concept for extending this flexible somatic base editing platform to other tissues and tumour types.

Keywords
base editing; breast cancer; CRISPR-Cas9; genetically engineered mouse models; intraductal injections

Introduction

Genetic sequencing studies defined a catalogue of somatic alterations in breast cancer [1]. However, deconvoluting the molecular complexity of breast tumours requires tractable and informative genetic models. Genetically engineered mouse models (GEMMs) represent the most sophisticated models of human breast cancer, as they simulate the stepwise progression of a healthy mammary cell to hyperplasia and invasive disease in the context of a native stromal compartment and in the presence of a functional immune system. However, the amount of resource and time required to derive new GEMM lines and to incorporate new mutant alleles within complex genotypes limits the experimental throughput.

In recent years, CRISPR-Cas9 genome editing has revolutionised gene function studies. The unprecedented ease with which endogenous loci can be perturbed with this method has opened a myriad of possibilities in terms of *in vivo* modelling of alterations observed in human malignancies. We previously showed that CRISPR-mediated somatic engineering of the mammary gland is feasible and effective using intraductal injection of lentivirally encoded sgRNAs in female Cas9 knock-in mice [2]. With this method, double-strand DNA breaks (DSB) can be generated *in situ* at a precise target location in the genome of mammary cells, and DNA repair processes such as non-homologous end joining (NHEJ) can result in the formation of insertions or deletions (indels), which may interrupt the open reading frame (ORF) and typically lead to gene disruption. This platform has proven instrumental in the assessment of the collaborative role of putative tumour suppressors in multiple breast cancer subtypes, including invasive lobular carcinoma (ILC) [3] and triple-negative breast cancer (TNBC) [4]. However, it is mostly applicable for probing the effects of complete loss of function of a candidate gene, whereas the most common disease-associated mutations seen in human breast cancer are point mutations [1], which can have more subtle consequences. Therefore, a way for rapidly installing precise mutations in the mouse mammary gland would provide a significant technological advance.

Base editing is a new genome editing technology which allows for the precise alteration of a DNA sequence without direct DSB formation (reviewed in [5]). The most characterised base editors, cytidine base editors (CBEs), are chimeric fusions composed of a nuclease-defective Cas9 tethering a cytidine

deaminase to specific DNA sequences to produce C-to-T transitions within defined windows of the protospacer.

In this study, we developed a knock-in mouse model for Cre-conditional expression of the BE3 cytidine base editor [6] in the mammary gland. We injected these mice with lentiviral vectors encoding one or multiple arrayed sgRNAs designed to install missense or nonsense mutations at one or multiple endogenous loci. This platform enabled rapid modelling of oncogenic variants and allelic series of oncogenes and tumour suppressors *in vivo*, and to test their contribution to tumourigenesis in a model of TNBC.

Results

Although BRCA1-associated TNBC is primarily a copy-number driven disease, mutations in TP53 and the PI3K/AKT pathway are, together with MYC copy-number variations, the most prominent aberrant events in these tumours [4]. We previously employed the *WapCre;Brca1*$^{F/F}$ *;Trp53*$^{F/F}$ *;Col1a1*$^{invCAG\text{-}Cas9/+}$ (WB1P-Cas9) mouse model of BRCA1-associated TNBC. In this model, mammary-specific expression of Cre induces inactivation of BRCA1 and p53 and concomitant expression of Cas9. We could use intraductal injection of Lenti-sgRNA-Myc lentiviral vectors in WB1P-Cas9 mice to test how disruption of specific genes (e.g., Pten or Rb1) collaborates with MYC overexpression in BRCA1-associated TNBC formation [4].

In order to model missense mutations rather than gene disruptions *in situ*, we generated a mouse model with conditional expression of the base editor BE3 in the mammary gland. The BE3 CBE is a hybrid protein that comprises the Streptococcus pyogenes Cas9 nickase (SpCas9D10A) fused with the rat APOBEC1 cytidine deaminase and a uracil glycosylase inhibitor (UGI) domain [6].Upon delivery of an sgRNA, the Cas9 moiety of BE3 engages with the genomic target site and positions the deaminase enzyme at its 5' end, where C-to-T transitions may be generated within a small 4–5 nucleotide window. *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ (WB1P-BE3) mice were generated using our previously established GEMM-ESC pipeline [7]. In brief, a Cre-conditional *invCAG-BE3* allele (**Appendix Figure S1A**) was introduced into the *Col1a1* locus of embryonic stem cells (ESCs) derived from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$ (WB1P) mice and chimeric mice were produced by blastocyst injection of the modified cells. High-quality male chimeras were

then back-crossed with *Brca1*$^{F/F}$;*Trp53*$^{F/F}$ females to generate the experimental cohort. In this WB1P-BE3 model, female mice spontaneously developed mammary tumours with a median latency of 195 days (n=17; **Appendix Figure S1B**), which is comparable to the previously reported latency of WB1P females (198 days) [4]. Similarly to WB1P tumours, WB1P-BE3 tumours were poorly differentiated carcinomas with a solid growth pattern, negative for oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) (**Figure EV1A**). To confirm that tumours from this new mouse model recapitulate the basal-like phenotype typical for WB1P tumours and for human BRCA1-associated breast cancer [4], we performed RNA sequencing on 6 WB1P-BE3 tumours and compared their expression profile to tumours from published mouse models of luminal (*WapCre;Cdh1*$^{F/F}$;*Pten*$^{F/F}$, WEP) and basal-like (*K14Cre;Brca1*$^{F/F}$;*Trp53*$^{F/F}$, KB1P; *WapCre;Brca1*$^{F/F}$;*Trp53*$^{F/F}$, WB1P; *WapCre;Brca1*$^{F/F}$;*Trp53*$^{F/F}$;*Col1a1*$^{invCAG\text{-}Myc/+}$, WB1P-Myc) breast cancer [4,8,9]. Unsupervised hierarchical clustering of gene expression profiles using a three-gene signature that distinguishes the PAM50 subtypes [10] and principal component analysis (PCA) of global gene expression confirmed that tumours from WB1P-BE3 mice retained a basal-like transcriptional identity (**Figure EV1B and C**).

We then cloned a lentiviral vector encoding an sgRNA targeting the third exon of *Akt1* in order to establish an oncogenic E17K missense mutation by base editing (*Akt1*^{E17K}). To validate this sgRNA, we transduced NIH3T3 cells expressing an optimized BE3 enzyme, FNLS [11], with Lenti-sg*Akt1*^{E17K} or a control Lenti-sgNT vector encoding a non-targeting sgRNA, and analysed targeted editing at the *Akt1* locus by Sanger sequencing 5 days after transduction. Cells transduced with Lenti-sg*Akt1*^{E17K} showed extensive target C-to-T conversion, leading to oncogenic AKT1 E17K mutations (**Figure 1A and B**), as well as bystander edits at nearby cytosines with variable efficiency (**Appendix Figure S2A**). As off-target base editing activity of CBEs has recently been reported [12,13], we performed whole-genome sequencing (WGS) of genomic DNA isolated from NIH3T3 cells with or without expression of the CBE and the sgRNAs, and performed genome-wide characterisation of off-target single-nucleotide variants (SNVs). As expected, the on-target edits could be readily detected at high allele frequencies in CBE-expressing cells transduced with Lenti-sg*Akt1*^{E17K}. While a limited number of additional SNVs could be detected, none of these off-target edits generated missense or nonsense mutations or altered essential splice sites (**Appendix Figure S3A**).

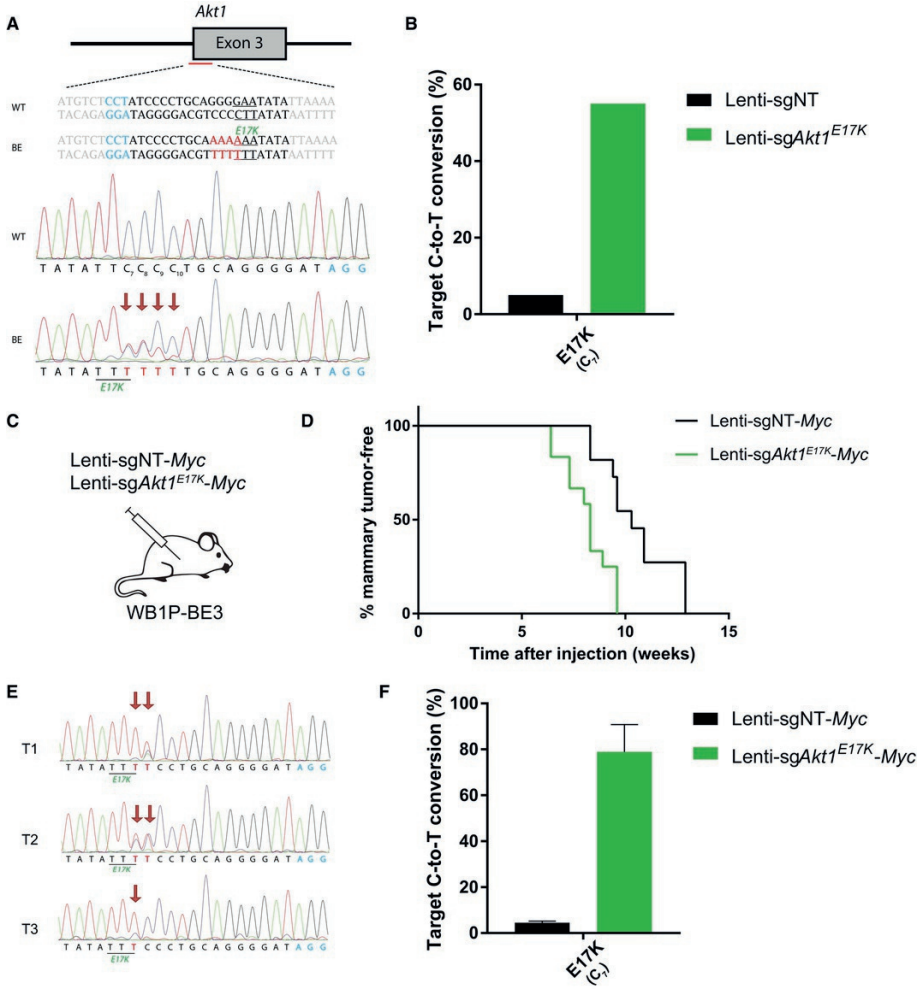


Figure 1: *In vivo* installation of oncogenic mutations by base editing in a model of triple-negative breast cancer

A) Sanger-sequencing chromatograms showing the target region of sg*Akt1*^{E17K} in wild-type (WT) and base edited (BE) cells. Arrowheads highlight cytosines of the protospacer that show base editing 5 days after transduction of BE3-expressing NIH3T3 cells with Lenti-sg*Akt1*^{E17K}. **B**) EditR [14] was used to calculate the frequency (%) of C-to-T conversion at C7 of the protospacer targeted by sg*Akt1*^{E17K} in BE3-expressing NIH3T3cells 5 days after transduction with the indicated sgRNA vectors. **C**) Overview of the intraductal injections performed in *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ (WB1P-BE3) females with high-titer lentiviruses encoding Myc cDNA and either a non-targeting (NT) sgRNA (Lenti-sgNT-Myc) or the sgRNA targeting *Akt1* (Lenti-sg*Akt1*^{E17K}-Myc). **D**) Kaplan–Meier curves showing mammary tumour-specific survival for the different models. WB1P-BE3 females injected with Lenti-sg*Akt1*^{E17K}-Myc (n=12) showed a reduced mammary tumour-specific survival compared to WB1P-BE3 female mice injected with Lenti-sgNT-Myc (n=11) vectors (58 days after injection vs. 72 days after injection, $P < 0.01$ by Mantel-Cox test). **E**) Sanger-sequencing chromatograms showing the target region of sg*Akt1*^{E17K} in three independent tumours from WB1P-BE3 females injected with Lenti-sg*Akt1*^{E17K}-Myc. Arrowheads highlight cytosines of the protospacer that show base editing. **F**) EditR was used to calculate the average frequency (%) of C-to-T conversion at C7 of the protospacer in tumours from WB1P-BE3 females injected with Lenti-sgNT-Myc or Lenti-sg*Akt1*^{E17K}-Myc. Data are plotted as mean + standard deviation (n=7).

To test the collaborative role of MYC overexpression and *Akt1*^{E17K} missense mutations *in vivo*, we generated lentiviral vectors encoding a Myc-overexpressing cassette together with the validated sg*Akt1*^{E17K} [4]. These vectors (Lenti-sgNT-Myc and Lenti-sg*Akt1*^{E17K}-Myc) were injected intraductally into WB1P-BE3 females (**Figure 1C**). As expected, all mice from both groups developed mammary tumours in the injected glands with 100% penetrance (**Figure 1D**). WB1P-BE3 mice injected with Lenti-sgNT-Myc developed mammary tumours with a median latency of 72 days after injection (n = 11), closely resembling latencies previously observed for WB1P-Cas9 mice injected with the same construct [4]. On the contrary, WB1P-BE3 mice injected with Lenti-sg*Akt1*^{E17K}-Myc developed tumours with a significantly shorter latency of 58 days (n = 12). Genomic DNA of mammary tumours from Lenti-sg*Akt1*^{E17K}-Myc injected WB1P-BE3 mice showed extensive editing of the target gene (**Figure 1E and F**), with greater than 78% average C-to-T conversion leading to activating *Akt1*^{E17K} missense mutations. Notably, bystander C-to-T editing and product purity at nearby cytosines of the protospacer were significantly lower, demonstrating positive selection specifically for oncogenic E17K mutations and not for other amino acid changes (**Appendix Figure S2B and C**). These results show that *in situ* base editing of the mammary gland enables modelling of defined point mutations within specific target genes.

We next tested whether this somatic platform could be used to generate an allelic series of missense mutations of an oncogene *in vivo*. The most frequent alterations observed in human BRCA1-associated TNBC, besides *TP53* alterations and *MYC* amplification, are *PIK3CA* missense variants [4,15]. We therefore designed multiple sgRNAs targeting *Pik3ca* and validated their ability to produce the hotspot E542K or E545K mutations (which are frequently observed in human tumours) or the much rarer E453K missense variant by base editing *in vitro* by Sanger sequencing and WGS (**Figure 2A and B, Appendix Figure S3B**). To test and compare the synergistic effect of MYC overexpression and *Pik3ca* missense mutations *in vivo*, we cloned Lenti-sg*Pik3ca*-Myc vectors encoding the specific sgRNAs targeting *Pik3ca*. The vectors were injected in WB1P-BE3 female mice (**Figure 2C**) and produced mammary tumours in all injected glands after variable latencies (**Figure 2D**). WB1P-BE3 females injected with Lenti-sg*Pik3ca*^{E542K}-Myc and Lenti-sg*Pik3ca*^{E545K}-Myc developed tumours significantly faster than Lenti-sgNT-Myc injected mice, with a median latency of 47 and 44 days after injection, respectively (n=10 and n=10). Notably, also mice injected with Lenti-sg*Pik3ca*^{E453K}-Myc developed tumours with a short median latency

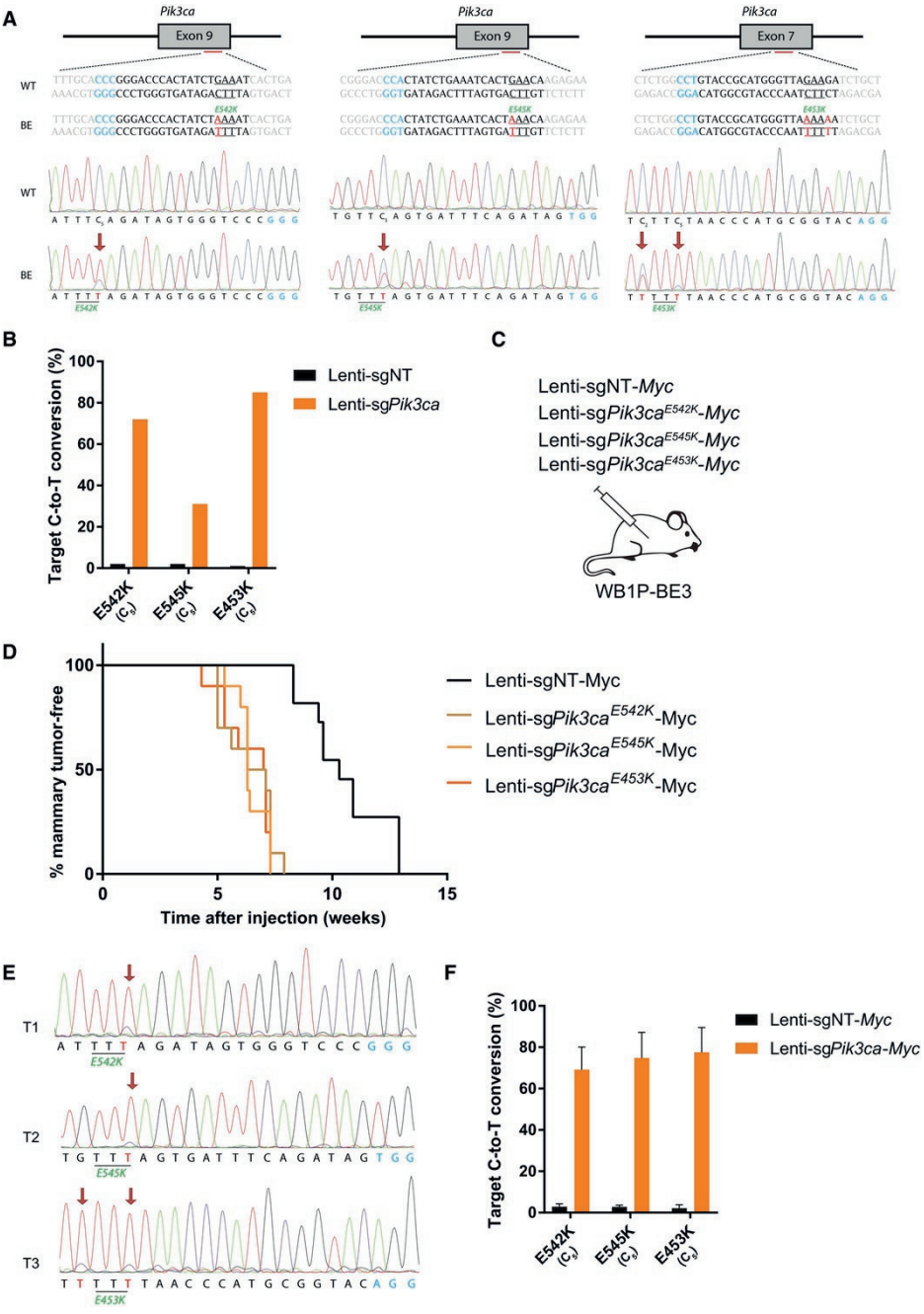

A
Pik3ca
Exon 9
Pik3ca
Exon 9
Pik3ca
Exon 7
WT
BE
E542K
E545K
E453K
B
Target C-to-T conversion (%)
Lenti-sgNT
Lenti-sgPik3ca
E542K (C9)
E545K (C9)
E453K (C9)
C
Lenti-sgNT-Myc
Lenti-sgPik3caE542K-Myc
Lenti-sgPik3caE545K-Myc
Lenti-sgPik3caE453K-Myc
WB1P-BE3
D
% mammary tumor-free
Time after injection (weeks)
Lenti-sgNT-Myc
Lenti-sgPik3caE542K-Myc
Lenti-sgPik3caE545K-Myc
Lenti-sgPik3caE453K-Myc
E
T1
T2
T3
F
Target C-to-T conversion (%)
Lenti-sgNT-Myc
Lenti-sgPik3ca-Myc

Figure 2: *In situ* base editing creates allelic series of oncogenic driver mutations

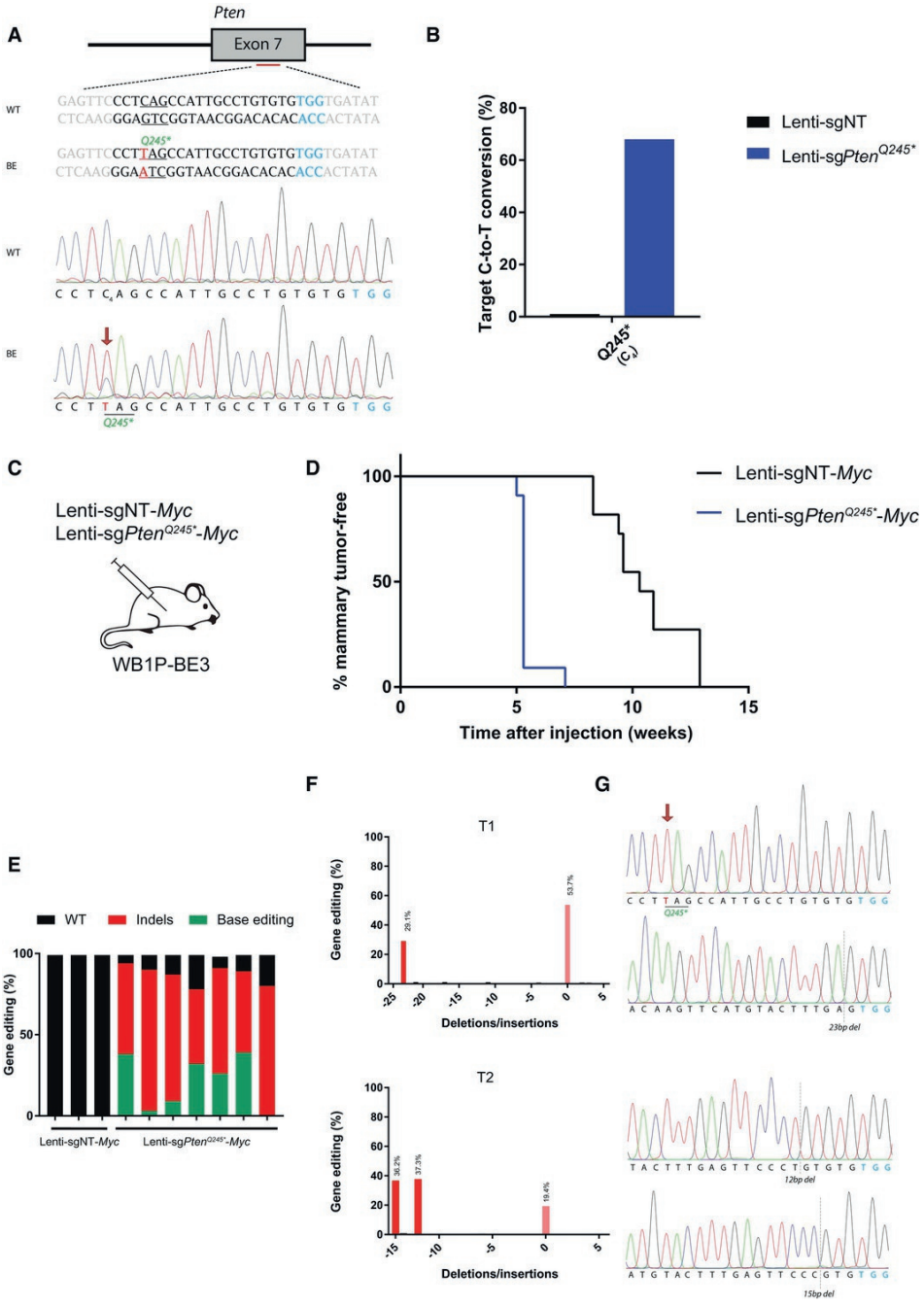
A) Sanger-sequencing chromatograms showing the target regions of sg*Pik3ca*^{E542K}, sg*Pik3ca*^{E545K}, and sg*Pik3ca*^{E453K} in wild-type (WT) and base edited (BE) cells. Arrowheads highlight cytosines of the protospacers that show base editing 5 days after transduction of BE3-expressing NIH3T3 cells with Lenti-sg*Pik3ca*^{E542K}, Lenti-sg*Pik3ca*^{E545K}, and Lenti-sg*Pik3ca*^{E453K}. **B**) EditR was used to calculate the frequency (%) of C-to-T conversion at the indicated target cytosines of the protospacers in BE3-expressing NIH3T3 cells 5 days after transduction with the indicated sgRNA vectors. **C**) Overview of the intraductal injections performed in WB1P-BE3 females with high-titer lentiviruses encoding Myc and either a non-targeting sgRNA (Lenti-sgNT-Myc)or the different sgRNAs targeting *Pik3ca* (Lenti-sg*Pik3ca*-Myc). **D**) Kaplan-Meier curves showing mammary tumour-specific survival for the different models. WB1P-BE3 females injected with Lenti-sg*Pik3ca*^{E542K}-Myc (n=10), Lenti-sg*Pik3ca*^{E545K}-Myc (n=10), and Lenti-sg*Pik3ca* ^{E453K}-Myc (n=10) showed a reduced mammary tumour-specific survival compared to WB1P-BE3 female mice injected with Lenti-sgNT-Myc (n=11) vectors (47, 44, and 49 days after injection, respectively vs. 72 days after injection, $P < 0.0001$ by Mantel-Cox test). **E**) Sanger-sequencing chromatograms showing the target region of sg*Pik3ca*^{E542K}, sg*Pik3ca*^{E545K}, and sg*Pik3ca*^{E453K} in three independent tumours from WB1P-BE3 females injected with the corresponding Lenti-sg*Pik3ca*-Myc vectors. Arrowheads highlight cytosines of the protospacer that show base editing. **F**) EditR was used to calculate the average frequency (%) of C-to-T conversion at the indicated target cytosines of the protospacers in tumours from WB1P-BE3 females injected with Lenti-sgNT-Myc or Lenti-sg*Pik3ca*^{E542K}-Myc, Lenti-sg*Pik3ca*^{E545K}-Myc and Lenti-sg*Pik3ca*^{E453K}-Myc. Data are plotted as mean + standard deviation (n=7).

of 49 days (n=10), under-scoring that the *Pik3ca*^{E453K} mutation, albeit less frequent than *Pik3ca*^{E542K} and *Pik3ca*^{E545K} in human tumours, has similar cooperative effects in this setting. By target sequencing of the tumours, we found average C-to-T editing to be 69, 75, and 78% for *Pik3ca*^{E542K}, *Pik3ca*^{E545K}, and *Pik3ca*^{E453K}, respectively (**Figures 2E and F, and EV2A**). As an additional control, we designed an sgRNA targeting intron 9 of the *Pik3ca* gene (sg*Pik3ca*^{intron}), immediately down-stream of the region targeted by sg*Pik3ca*^{E542K} and sg*Pik3ca*^{E545K}. As this region is reasonably distant from the exon-intron junction, we expected base conversions at this site to have neutral consequences on PIK3CA expression and activity. We validated the capability of sg*Pik3ca*^{intron} to produce specific C-to-T conversions at the target *site in vitro* by Sanger sequencing (**Figure EV2B**). We then cloned a Lenti-sg*Pik3ca*^{intron}-Myc construct which we injected intraductally into WB1P-BE3 mice. These mice developed tumours after a median latency of 67 days (n=9), comparable to the tumour latency of WB1P-BE3 mice injected with Lenti-sgNT-Myc and significantly later than WB1P-BE3 mice injected with the codon-targeting Lenti-sg*Pik3ca*-Myc vectors (**Figure EV2C**). These data further support that the shortened tumour latency of the latter is due to the specific mutations installed by base editing.

The high C-to-T rates achieved *in vivo* with Lenti-sg*Akt1*-Myc and Lenti-sg*Pik3ca*-Myc vectors indicate that continuous editing during tumour progression could saturate base conversion at the target site in both copies of *Akt1* or *Pik3ca*. Therefore, we next tested whether we could apply *in situ* base editing for bi-allelic inactivation of a tumour suppressor gene. We designed an sgRNA targeting the tumour suppressor *Pten*, and we validated the capability of Lenti-sg*Pten*^{Q245*} to create nonsense editing *in vitro* by target sequencing and WGS (**Figure 3A and B, Appendix Figure S3C**). We then injected WB1P-BE3 mice with Lenti-sg*Pten*^{Q245*}-Myc vectors (n=11) with the goal of overexpressing MYC and inactivating *Pten*, and observed accelerated TNBC formation in these mice compared with WB1P-BE3 mice injected with Lenti-sgNT-Myc (**Figure 3C and D**). The average latency (37 days after injection) was comparable to the mammary tumour-free survival of WB1P-Cas9 mice injected with the same Lenti-sg*Pten*-Myc construct [4], indicating that in both cases loss of function of *Pten* was collaborating with MYC overexpression in BRCA1-associated mammary tumourigenesis. On the contrary, WB1P mice injected with Lenti-sg*Pten*^{Q245*}-Myc (n=11) developed TNBC with a median latency of 69 days, comparable to control tumours, further confirming that only the combined expression of BE3 and sg*Pten*^{Q245*} is responsible for the short tumour latency in WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc

Figure 3: *In vivo* nonsense editing of Pten

A) Sanger-sequencing chromatograms showing the target region of sg*Pten*^{Q245*} in wild-type (WT) and base edited (BE) cells. Arrowheads highlight cytosines of the protospacer that show base editing 5 days after transduction of BE3-expressing NIH3T3 cells with Lenti-sg*Pten*^{Q245*}. **B**) EditR was used to calculate the frequency (%) of C-to-T conversion at C4 of the protospacer targeted by sg*Pten*^{Q245*} in BE3-expressing NIH3T3 cells 5 days after transduction with the indicated sgRNA vectors. **C**) Overview of the intraductal injections performed in WB1P-BE3 females with high-titer lentiviruses encoding Myc and either a non-targeting sgRNA (Lenti-sgNT-Myc) or the sgRNA targeting *Pten* (Lenti-sg*Pten*^{Q245*}-Myc). **D**) Kaplan–Meier curves showing mammary tumour-specific survival for the different models. WB1P-BE3 females injected with Lenti-sg*Pten*^{Q245*}-Myc (n=11) showed a reduced mammary tumour-specific survival compared to WB1P-BE3 female mice injected with Lenti-sgNT-Myc (n=11) vectors (37 days after injection vs. 72 days after injection, $P < 0.0001$ by Mantel-Cox test). **E**) BE Analyzer [16] was used to assess from next-generation sequencing data the fraction of wild-type *Pten* alleles, base edited alleles, or alleles with insertions/deletions (indels) in tumours from WB1P-BE3 animals injected with Lenti-sgNT-Myc or Lenti-sg*Pten*^{Q245*}-Myc. **F**) TIDE analysis showing the spectrum of indels of the targeted *Pten* alleles in two independent representative tumours from WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc. **G**) For the two tumours shown in (F), Sanger-sequencing chromatograms showing the target region of sg*Pten*^{Q245*} (PCR products were subcloned for clarity). Arrowheads highlight cytosines of the protospacer that show base editing. In the lower example, the gene was inactivated by indels at both alleles, while in the upper one by Q245* base editing in one allele and a deletion at the second copy of the gene.


A
Pten
Exon 7
WT
BE
Q245*
B
Target C-to-T conversion (%)
Lenti-sgNT
Lenti-sgPten Q245*
Q245* (C)
C
Lenti-sgNT-Myc
Lenti-sgPten Q245*-Myc
WB1P-BE3
D
% mammary tumor-free
Time after injection (weeks)
Lenti-sgNT-Myc
Lenti-sgPten Q245*-Myc
E
WT
Indels
Base editing
Gene editing (%)
Lenti-sgNT-Myc
Lenti-sgPten Q245*-Myc
F
T1
T2
Gene editing (%)
Deletions/insertions
G
23bp del
12bp del
15bp del

(**Figure EV3A**). Indeed, tumours from the latter group showed decreased PTEN levels and displayed activation of the PI3K/AKT downstream signalling pathway as visualised by immunoblot and immunohistochemical analysis of PTEN, phospho-Akt^{Ser473}, and phospho-S6$^{Ser235/236}$ expression (**Figures EV3B and C, and EV4**).

To characterise the phenotypes of the base edited mammary tumours described so far in more detail, we performed RNA sequencing on a panel of 29 additional tumours from WB1P-BE3 mice injected with different Lenti-sgRNA-Myc vectors, and compared their expression profiles to those of spontaneous WB1P-BE3 tumours. The tumours from the somatic models clustered together based on gene expression, but separated from spontaneous WB1P-BE3 tumours (**Figure EV5A**). Nonetheless, unsupervised hierarchical clustering of gene expression profiles using a three-gene signature that distinguishes the PAM50 subtypes [10] and PCA analysis of global gene expression confirmed that all tumours from the somatic models retained a basal-like transcriptional identity (**Figure EV5B and C**). Histopathological analysis confirmed that they were all comparable to WB1P-BE3 tumours in terms of morphology and expression of ER, PR, HER2, E-cadherin, vimentin, keratin 8, and keratin 14, and despite higher MYC expression, they showed similar Ki-67 stainings (**Figures EV4 and EV5D, Appendix Figure S4**). Elevated phospho-S6$^{Ser235/236}$ expression was obvious only in tumours from WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc, and to a lower extent in tumours from WB1P-BE3 mice injected with Lenti-sg*Akt1*^{E17K}-Myc, but not in tumours from WB1P-BE3 mice injected with Lenti-sg*Pik3ca*-Myc vectors (**Figure EV4**). In accordance with this, gene set enrichment analysis (GSEA) indicated activation of the mTORC1 signalling in tumours from WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc and Lenti-sg*Akt1*^{E17K}-Myc, but not in tumours from WB1P-BE3 mice injected with Lenti-sg*Pik3ca*-Myc and Lenti-sgNT-Myc vectors (**Appendix Figure S5**).

Notably, target sequencing of *Pten* in tumours from WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc showed that in these specimens the gene was inactivated either by frame-shifting indels at both *Pten* alleles or by Q245* base edits in one allele and indels at the second copy of the gene (**Figure 3E–G, Appendix Figure S6A**). It was previously shown that BE3 can yield low but detectable unintended indels instead of base alterations *in vitro* [6]. However,

as we did not observe evident by-product indels in tumours somatically base edited with sg*Akt1* or sg*Pik3ca*, we reasoned that they might only become apparent in our somatic model when targeting a tumour suppressor gene like *Pten*, in which gene disruption by truncation is likely selected to the same extent as gene inactivation by nonsense mutation. To investigate this further, we designed and validated an sgRNA capable to install a Q97* ocher mutation in *Trp53 in vitro* by Sanger sequencing and WGS (**Figure 4A and B, Appendix Figure S3D**). Then, we generated *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ mice with heterozygous *Trp53*^{F} floxed alleles and intraductally injected them with Lenti-sgNT-Myc or Lenti-sg*Trp53*^{Q97*}-Myc (**Figure 4C**). Moreover, to test the feasibility of multiplexed *in vivo* base editing, we also injected these mice with a tandem Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc vector that harbours two arrayed sgRNA cassettes, to simultaneously introduce the missense *Pik3ca*^{E545K} mutation and inactivate the residual wild-type copy of *Trp53*. *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ females injected with Lenti-sgNT-Myc (n=11) did not develop any palpable tumours during the 150 days of observation period (**Figure 4D**). In contrast, mice injected with Lenti-sg*Trp53*^{Q97*}-Myc and Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc developed TNBC tumours after a median latency of 101 and 76 days, respectively (n=5 and n=6). Most of the tumours from mice injected with Lenti-sg*Trp53*^{Q97*}-Myc and Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc displayed the targeted *Trp53*^{Q97*} mutation achieved by C-to-T base editing at C8 of the protospacer (**Figure 4E–G, Appendix Figure S6B**), always together with a collateral edit at a nearby cytosine (C6). Also in this case however, in some tumours the *Trp53* allele displayed a frame-shifting indel within the protospacer instead. Notably, target sequencing of *Trp53* showed that bystander editing at C6 was still present in tumours with indels, suggesting that an initially base edited allele was re-targeted by the protracted activity of the CRISPR machinery, producing a DSB which was then resolved by indel-prone end-joining processes (**Figure 4G**). On the contrary, target sequencing of the *Pik3ca* gene confirmed that tumours induced by the tandem Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc vector displayed almost exclusively E545K base edits (79% average C-to-T editing), although bystander indels could be detected in a minor allele fraction upon deep sequencing (**Appendix Figure S6C and D**).

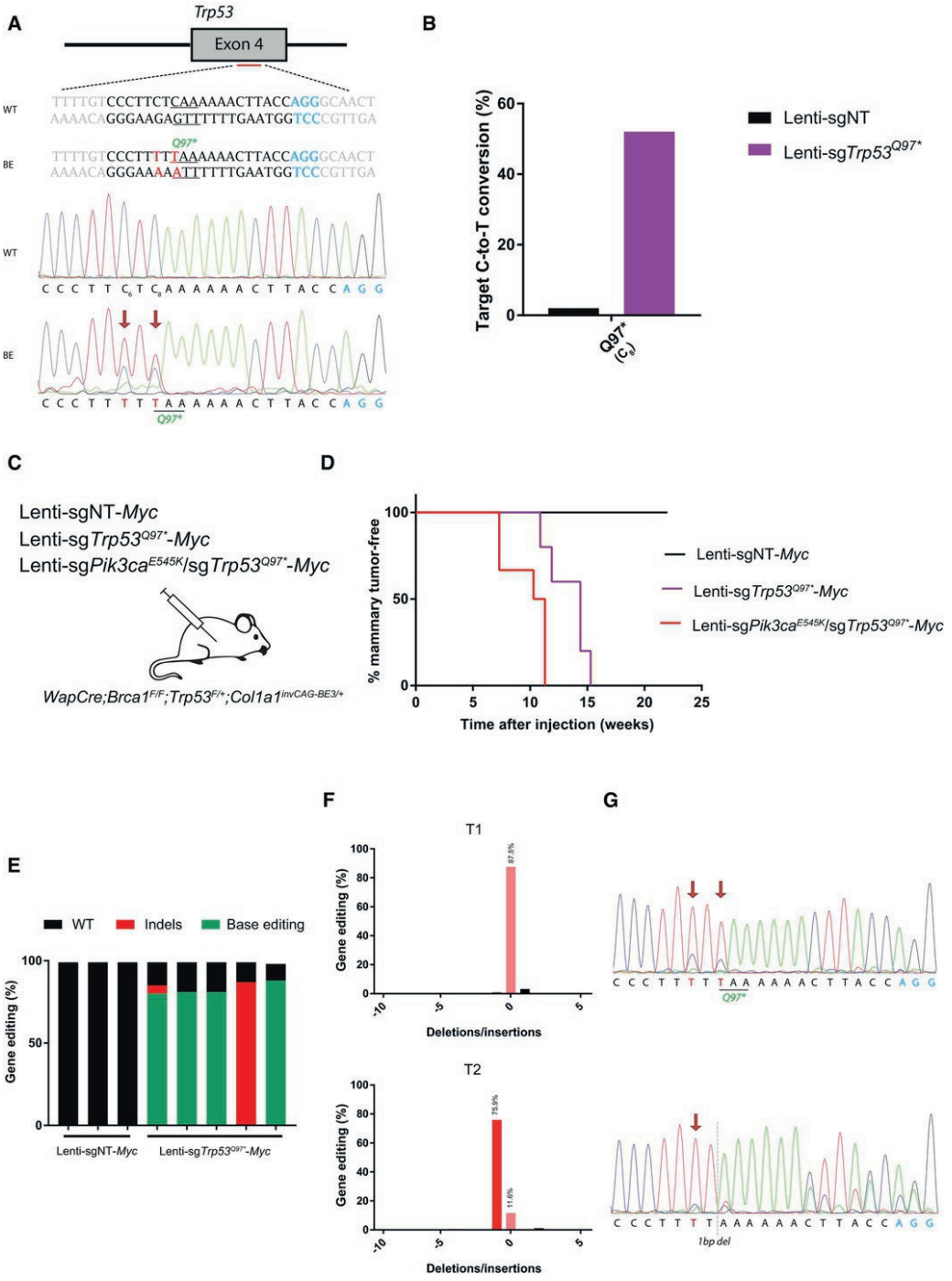

A
Trp53
Exon 4
WT
BE
Q97*
B
Target C-to-T conversion (%)
Lenti-sgNT
Lenti-sgTrp53Q97*
Q97* (C8)
C
Lenti-sgNT-Myc
Lenti-sgTrp53Q97*-Myc
Lenti-sgPik3caE545K/sgTrp53Q97*-Myc
WapCre;Brca1F/F;Trp53F/+;Col1a1invCAG-BE3/+
D
% mammary tumor-free
Time after injection (weeks)
Lenti-sgNT-Myc
Lenti-sgTrp53Q97*-Myc
Lenti-sgPik3caE545K/sgTrp53Q97*-Myc
E
WT
Indels
Base editing
Gene editing (%)
Lenti-sgNT-Myc
Lenti-sgTrp53Q97*-Myc
F
T1
T2
Gene editing (%)
Deletions/insertions
87.5%
75.9%
11.6%
G
1bp del

Figure 4: Multiplexed *in vivo* base editing

A) Sanger-sequencing chromatograms showing the target region of sg*Trp53*^{Q97*} in wild-type (WT) and base edited (BE) cells. Arrowheads highlight cytosines of the protospacer that show base editing 5 days after transduction of BE3-expressing NIH3T3 cells with Lenti-sg*Trp53*^{Q97*}. **B**) EditR was used to calculate the frequency (%) of C-to-T conversion at C8 of the protospacer targeted by sg*Trp53*^{Q97*} in BE3-expressing NIH3T3 cells 5 days after transduction with the indicated sgRNA vectors. **C**) Overview of the intraductal injections performed in *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ (Trp53^{F}-het WB1P-BE3) females with high-titer lentiviruses encoding Myc and either a non-targeting sgRNA (Lenti-sgNT-Myc), the sgRNA targeting *Trp53* (Lenti-sg*Trp53*^{Q97*}-Myc), or two arrayed sgRNA cassettes encoding sg*Pik3ca*^{E545K} and sg*Trp53*^{Q97*} (Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc). **D**) Kaplan-Meier curves showing mammary tumour-specific survival for the different models. *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ females injected with Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc (n=6) showed a reduced mammary tumour-specific survival compared to animals injected with Lenti-sg*Trp53*^{Q97*}-Myc (n=5) vectors (76 days after injection vs. 101 days after injection, $P < 0.05$ by Mantel-Cox test). Females injected with Lenti-sgNT-Myc (n=11) did not develop any palpable tumours during the 150 days of observation period. **E**) BE Analyzer was used to assess from next-generation sequencing data the fraction of wild-type *Trp53* alleles, base edited alleles, or alleles with indels in tumours from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ animals injected with Lenti-sg*Trp53*^{Q97*}-Myc. Tumours from WB1P-BE3 animals injected with Lenti-sgNT-Myc mice were used as control. **F**) TIDE analysis showing the spectrum of indels of the targeted *Trp53* alleles in two independent representative tumours from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ mice injected with Lenti-sg*Trp53*^{Q97*}-Myc. **G**) For the two tumours shown in (F), Sanger-sequencing chromatograms showing the target region of sg*Trp53*^{Q97*}. Arrowheads highlight cytosines of the protospacer that show base editing. In the lower example, the gene was inactivated by a deletion, while in the upper one by Q97* base editing. Of note, the allele with the indel also displays base editing at C6 of the protospacer.

Discussion

Most human cancers are predominantly characterised by missense mutations. Here, we show that somatic base editing is feasible and effective at installing defined missense and nonsense mutations at endogenous loci in a mouse model of TNBC. The possibility to rapidly engineer breast cancer-associated point mutations *in situ* allows us to recapitulate gain-of-function mutations in known and putative oncogenes in preclinical models, and to evaluate the relative effect size of each genetic perturbation within an allelic series. Somatic base conversion for cancer modelling has previously been achieved *in situ* by hydrodynamic injection of plasmids encoding BE3 and an sgRNA in the mouse tail vein, which led to oncogenic C-to-T editing at the b-catenin gene in the adult liver [11]. This approach is not applicable in the mammary gland, as we previously observed that *de novo* expression of Cas9 in adult mice elicits strong immune infiltration in this compartment, which could be circumvented by expressing the bacterial endonuclease from a conditional knock-in allele [2]. Following the same paradigm, we report here the generation of a knock-in mouse model harbouring a Cre-conditional BE3 allele, and its

validation as a flexible and multiplexable platform for *in situ* base editing of the mammary gland upon intraductal delivery of sgRNA-encoding vectors. Using this system, we validated loss of function of *PTEN* and activation of *AKT1* and *PIK3CA* as *bona fide* drivers of BRCA1-associated tumourigenesis. Moreover, the possibility to rapidly derive cohorts of tumours engineered with defined mutations allowed us to evaluate the effect on tumourigenesis of different allelic variants of *Pik3ca*. This pipeline can also be used to test the effects of clinically relevant missense mutants on therapy response by orthotopic transplantation of tumour fragments or tumour-derived organoids into syngeneic mice [17,18].

A potential limitation of our system comes from the protracted expression of the CBE in the mammary gland of WB1P-BE3 mice, which often saturates base conversion at both alleles of an oncogene. This sustained expression could also increase off-target mutation rates and unintended by-product indel formation. Indeed, when targeting *Pten* and *Trp53* with the goal to install premature stop codons, we found a subset of the tumour suppressor alleles displayed gene inactivation by indels instead of nonsense edits. Possible solutions to minimise this downside could entail strategies to control editing dynamics using inducible or self-inactivating editors, selection against DSB formation with CBEs that encode DNA end-binding Gam proteins [19] or switch to systems based on nuclease-dead Cas9 rather than nickase. It is worth mentioning, however, that in cases where the effect of a missense mutation in a candidate cancer gene is unknown, the product promiscuity of our somatic platform could shed light on whether the functional consequence of the mutation is likely a gain of function or loss of function. Two recent papers have shown that the CBE off-target mutation rate is higher than previously anticipated [12,13]. Although off-target activity of CBEs can be detrimental for therapeutic applications, it is much less of an issue for tumour acceleration studies in mouse models, in which random mutations collaborate with the genetically engineered mutations in driving tumourigenesis. This is particularly true in our TNBC mouse model, in which mammary tumourigenesis is induced by engineered loss of BRCA1 and p53, which results in loss of homologous recombination (HR) repair, genomic instability and a mutator phenotype. Still, while non-sequence-dependent off-targets can be experimentally controlled with neutral sgRNAs and biological replicates, sgRNA-dependent off-targets should be scrutinised on a case-by-case basis, preferentially by WGS.

Even with high-fidelity editors, a current limitation of CRISPR-mediated base editing is that not all missense variants can be modelled with the same enzymes. For example, the most prevalent *PIK3CA* mutation, H1047R, requires a G-to-A conversion that cannot be produced with CBEs, but only with recently described adenine base editors (ABEs) [20]. Moreover, to enable efficient base editing, a protospacer adjacent motif (PAM) needs to be present and appropriately distanced from a target base. Finally, some mutations that require C-to-non-T editing are less favoured, especially with UGI-encoding editors. However, the base editing field is rapidly evolving to expand the range of targetable codons. Recently, base editors encoding alternative Cas9 orthologues or engineered SpCas9 variants that recognise a broader range of PAM shave been optimised [21-24]. In parallel, CBEs have been developed with reduced or expanded width of the editing window, to minimise bystander editing at non-target cytosines or to enlarge the repertoire of targetable bases, respectively [11,25-28]. Finally, base editors that efficiently convert target cytosines to a mixture of the other three bases have also been established [29] and might be particularly appealing for localised sequence diversification and mutagenesis *in vivo*. In general, as the catalogue of base editors with specific properties continues to expand, it may be relevant to develop knock-in mice with conditional expression of additional base editing enzymes.

In conclusion, our *in vivo* base editor model offers novel opportunities for fast-track generation of somatic GEMMs of breast cancer. The conditional BE3 allele allows *in vivo* characterisation of point mutations at a defined endogenous locus to assess their role in initiating or accelerating tumour formation in the mammary gland, alone or in combination with other conditional alleles. While we focused on TNBC in this study, the applicability of this strategy could be extended to other organs and tumour types by inter-crossing BE3 mice with different Cre-conditional mouse models.

Materials and Methods

sgRNA design

The sgRNAs for base editing were designed using Benchling (https://benchling.com). sg*Akt1*^{E17K}: TATATTCCCCTGCAGGGGAT; sg*Pik3ca*^{E545K}: TGTTCAGTGATTTCAGATAG; sg*Pik3ca*^{E542K}: ATTTCAGATAGTGGGTCCC; sg*Pik3ca*^{E453K}: TCTTCTAACCCATGCGGTAC; sg*Pik3ca*^{intron}: CTCTCAA-

GGCTGAAGGCCG; sg*Pten*^{Q245*}: CCTCAGCCATTGCCTGTGTG; sg*Trp53*^{Q97*}: CCCTTCTCAAAAAACTTACC.

Lentiviral vectors

The sgRNAs were cloned as described [30] into Lenti-U6-tdTomato-P2A-BlasR vectors (Lenti-sgRNA [11], Addgene plasmid #110854) or pGIN backbones [31]. All vectors were validated by Sanger sequencing. The pGINLenti-sgNT-Myc vector, encoding Myc cDNA, and a non-targeting sgRNA (TGATTGGGGGTCGTTCGCCA) were described before [4]. For cloning of other Lenti-sgRNA-Myc vectors, XbaI and XhoI were used to extract a Myc-encoding fragment from Lenti-sgNT-Myc, which was inserted in the XbaI-XhoI digested backbones of the pGIN vectors encoding the different sgRNAs. For cloning of the Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc tandem vector, a fragment encoding sg*Pik3ca*^{E545K} was amplified by PCR from Lenti-sg*Pik3ca*^{E545K}-Myc using XbaI-containing primers and cloned in the XbaI digested backbone of Lenti-sg*Trp53*^{Q97*}-Myc. pLenti-FNLS-P2A-Puro was a gift from Lukas Dow [11] (Addgene plasmid #110841). Concentrated stocks of VSV-G pseudotyped lentivirus were produced by transient co-transfection of four plasmids in 293T as previously described [32]. Lentiviral titers were determined using the qPCR lentivirus titration kit from Abm (LV900).

Cell culture

293T cells for lentiviral production and NIH3T3 cells were cultured in Iscove's medium (Invitrogen Life Technologies) containing 10% FBS, 100 IU ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin. All transductions were performed by adding diluted viral supernatant to the cells in the presence of 8 µg ml^{-1} polybrene (Sigma). For testing of sgRNA activity *in vitro*, NIH3T3 cells were first transduced with pLenti-FNLS-P2A-Puro, and after 3 days of 2 µg ml^{-1} puromycin selection, they were re-transduced with the different Lenti-sgRNA vectors and selected for 4 days with 4 µg ml^{-1} blasticidin. Harvesting of cells for genomic DNA isolation was performed 5 days after transduction with the Lenti-sgRNA vectors.

PCRs, Sanger sequencing, and EditR analyses

Genomic DNA from frozen cell pellets was isolated using the Gentra Puregene genomic DNA isolation kit from Qiagen. For Sanger sequencing, amplification of base edited targets was performed with specific primers spanning the target sites (FW_Akt1:CCTGCGTATGGCTGATGTTG; RV_Akt1: CCCGCATGGCTAAGACACTT; FW_Pik3ca_1: AGTGGAGTGTAGGAAGAGCCT; RV_Pik3ca_1: ACAGGAAGAA-

GGTCCCTCGG; FW_Pik3ca_2: ACCCTAGTGTCCGGGAAAATG; RV_Pik3ca_2: AGAGCTCAACAGTAGCCACAC;FW_Pten: TGTATTTAACCACACAGATCCTCA; RV_Pten: AACAAACTAAGGGTCGGGGC; FW_Trp53: CTTTGGTGTTGGGCTGG-TAG;RV_Trp53: GGGCAAAACTAAACTCTGAGGC) and 1 µg DNA template using the Q5 high-fidelity PCR kit from NEB. Amplicons were sequenced using the FW primer, and CRISPR/Cas9-induced base edits were quantified as described with EditR [14] (https://moriaritylab.shinyapps.io/editr_v10). Untransduced cells were taken along as a control in each amplification.

Mouse studies

pCMV-BE3 was a gift from David Liu (Addgene plasmid #73021). BE3 cDNA was sequence-verified and inserted as FseI-NotI fragments into the Frt-invCag-IRES-Luc shuttle vector [7], resulting in *Frt-invCag-BE3*. Flp-mediated knock-in of the shuttle vector in the *WapCre;Brca1$^{F/F}$;Trp53$^{F/F}$;Col1a1-frt* GEMM-ESC was performed as described [7]. Chimeric animals were crossed with *Brca1$^{F/F}$;Trp53$^{F/F}$* mice to generate the experimental cohorts. WapCre, *Brca1$^{F/F}$*, *Trp53$^{F/F}$*, and knock-in alleles were detected using PCR as described [7,8,33]. Intraductal injections were performed as described [2,34]. Lentiviral titres ranging from 2 to 20×10^{8} TU ml^{-1} were used. Animal experiments were approved by the Animal Ethics Committee of the Netherlands Cancer Institute. Mice were bred and maintained in accordance with institutional, national and European guidelines for Animal Care and Use.

Statistical analysis

Log-rank Mantel-Cox test was used for the following figures: Figures 1D, 2D, 3D, 4D, EV2C, and EV3A.

Immunoblotting

Protein lysates were made using lysis buffer (20 mM Tris pH 8.0,300 mM NaCl, 2% NP-40, 20% glycerol, 10 mM EDTA) complemented with protease inhibitors (Roche) and quantified using the BCA Protein Assay Kit (Pierce). Protein lysates were loaded onto a 4–12% Bis-Tris gradient gel (Invitrogen) and transferred on a nitro-cellulose membrane (Bio-Rad) in transfer buffer (38 mM glycine, 5 mM TRIS, and 0.01% SDS in PBS-T (0.5% Tween-20)). Membranes were blocked in 5% w/v bovine serum albumin (BSA) in PBS-T after which they were stained for 2 h at room temperature using the primary antibodies anti-AKT1 (1:1,000, Cell Signaling Technology (CST) 2938), anti-phospho-AKT1^{Ser473} (1:2,000, CST4060), anti-p44/42 MAPK (1:1,000, CST 4695), anti-phospho-p44/42MAPK ERK1/ERK2$^{Thr202/Tyr204}$ (1:2,000 CST 9101),

anti-S6 (1:1,000,CST 2217), anti-phospho-S6$^{Ser235/Ser236}$ (1:2,000, CST 2211), anti-PTEN (1:1,000 CST 9188), and anti-b-actin (1:50,000, Sigma A5441) in 5% w/v BSA in PBS-T. Membranes were washed three times with1% BSA in PBS-T and incubated for 1 h with an HRP-conjugated secondary antibody (1:2,000, DAKO). Stained membranes were washed three times in 1% BSA in PBS-T and developed using Pierce ECL Western Blotting Substrate (Thermo Scientific).

Histology and immunohistochemistry

Tissues were formalin-fixed overnight and paraffin-embedded by routine procedures. Hematoxylin and eosin staining was performed as described [35]. Immunohistochemical stainings were processed as described [35,36]. For ER, PR, and phospho-S6$^{Ser235/Ser236}$, primary mouse anti-body anti-ER (Santa Cruz sc-542), anti-PR (Thermo Scientific RM-9102), and anti-phospho-S6$^{Ser235/Ser236}$ (CST 2211) were used. For HER2, E-cadherin, vimentin, keratin 14, Myc and Ki-67, primary rabbit antibody anti-NEU (Santa Cruz sc-284), anti E-cadherin (CST3195), anti-vimentin (CST 5741), anti-cytokeratin 14 (Abcam ab181595), anti-MYC (Abcam ab32072), and anti-Ki-67 (Abcam ab15580) were used. For keratin 8, primary rat antibody anti-cytokeratin 8 (University of Iowa TROMA-1) was used. All slides were digitally processed using the Aperio ScanScope (Aperio, Vista, CA, USA) and captured using ImageScope software version 12.0.0 (Aperio).

Deep target sequencing of tumour fragments

Frozen tumour pieces were lysed overnight in lysis buffer (100 mMTris–HCl, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, and 100 µg ml^{-1} Proteinase-K), and genomic DNA was purified with standard phenol-chloroform extraction. For deep sequencing, amplification of base edited targets was performed with specific primers spanning the target sites and including the Phased PE adapter sequence (FW_PE-Pten: ACACTCTTTCCCTACACGACGCTCT-TCCGATCTTGTGGTCTGCCAGCTAAAGG; RV_PE-Pten: CGGTCTCGG-CATTCCTGCTGAACCGCTCTTCCGATCTCCACAGAAATGAAGAGTCTGCC; FW_PE-Trp53: ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTTTGAA-GGCCCAAGTGAAGC; RV_PE-Trp53: CGGTCTCGGCATTCCTGCTGAAC-CGCTCTTCCGATCTAGGCATTGAAAGGTCACACGA; FW_PE-Pik3ca: ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCACCAGTTTGCT-TTTTCAAAT; RV_PE-Pik3ca: CGGTCTCGGCATTCCTGCTGAACCGCTCTTC-CGATCTGACAGGAAGAAGGTCCCTCG) using Platinum Taq DNA Polymerase High Fidelity (ThermoFisher). Indexed libraries were sequenced using Illumina MiSeq technologies (Paired End 250 bp runs spiked with 50% PhiX). CRISPR/

Cas9-induced base edits and indels were quantified as described with BE Analyzer [16] (http://www.rgenome.net/be-analyzer). For Sanger sequencing, amplification of base edited targets was performed similarly as for cells. Amplicons were sequenced using the FW primer and CRISPR/Cas9-induced base edits and indels were quantified as described with EditR and TIDE [14,37] (https://moriaritylab.shinyapps.io/editr_v10; http://tide.nki.nl). Untransduced cells were taken along as a control in each amplification.

Generation and analysis of RNA sequencing data

The mRNA library was generated using Illumina TruSeq Stranded mRNA Library Prep Kit and sequenced with 65 base single reads on HiSeq 2500. The sequencing reads were first trimmed using Cutadapt (v.1.13) to remove any residual adapter sequences and filter the short reads smaller than 20 bp after trimming the adapter sequences. The trimmed reads were then mapped to the reference genome (Ensembl GRCm38) using STAR aligner (v.2.5.2b) [38]. The aligned reads were quantified using feature Counts (v. 1.5.2) [39] based on the gene annotation from Ensembl GRCm38 version 89. The raw gene read counts were normalised by TMM normalisation using edgeR [40], and count per million (CPM) values were computed using limma-voom [41]. Genes with CPM < 1 across the entire samples were excluded for downstream analysis to reduce the false positives that can derive from lowly expressed genes. The RNA sequencing data for basal (KB1P, WB1P, and WB1P-Myc) and luminal (WEP) tumours were obtained from previous studies from our group [4]. For integration with our new dataset, we used the raw read counts that were derived from the same pipeline. The raw read counts for the new and previous data-sets were then normalised together using TMM normalisation, and CPM values were computed using edgeR and limma-voom, as described above. Genes with CPM < 1 across the entire samples were excluded for downstream analysis. Gene set enrichment analysis was performed using fgsea with the gene set “MTORC1_SIGNALING” in the MSigDB Hallmark gene set collection [42]. Moderated t-statistics from limma-voom [41] was used to rank the genes, and the permutation for each gene set was conducted 10,000 times to obtain an empirical null distribution.

Generation and analysis of whole-genome sequencing data

Whole-genome sequencing libraries were prepared using standard protocols for the Illumina X10 platform. The resulting sequence was aligned using bwa-mem (v.0.7.17) to the reference mouse GRCm38 assembly, and PCR duplicates were marked using bam streaming-mark duplicates in biobambam

(v.2.0.79). The total mapped coverage varied from 24× to 44×, with a median of 37×. To identify off-target edits in cell lines with a targeting sgRNA, variant calling was performed using cgpCaVEManWrapper (v1.13.14) [43] and cgpPindel (v3.3.0) [44], using a control (with no sgRNA) as the reference sample. The raw CaVEMan and P indel calls were merged, and the bcftools [45] SnpGap filter was used to remove SNVs that were within 15 bp of an indel, as these are likely false-positive variants that are a result of mismatches from read alignment issues. Further false-positive variants were further filtered by selecting the "pass" calls tagged from the cgpCa-VEMan and cgpPindel default filtering, excluding calls (unfiltered) found in a second control (cells transduced with a non-targeting sgRNA), indels in simple repeats, and variants in common laboratory mouse strains from the Mouse Genomes Project (release version 6) [46]. To remove additional false-positive calls, indel calls that fell inside of any type of repeat were excluded, unless the whole repeat was deleted. Indel calls with at least 1 alternate allele in the reference sample (no sgRNA) were excluded. The following cgpCaVEMan filters were applied to SNVs: CLPM = 0 (no soft-clipped bases in the reads with the variant) and ASMD ≥ 130 (median alignment score of the reads with the variant). Variant calls with either allele frequency (AF) < 0.1, less than 5 variant-supporting reads (minimum base quality score 30), or at least 1 alternative allele in either control samples (minimum base quality score 30) were also considered false positives. SNV and indel calls at sites with less than 10× total coverage in either the reference or query samples were also excluded. After visual inspection of the read alignments for remaining indel calls and comparison to the same regions in the controls, it was determined that these calls were likely false-positive indels. To determine which SNVs are potentially deleterious, VAGrENT [47] was used to identify missense mutations, nonsense mutations, and variants altering essential splice sites.

Data availability

The RNA-Seq data produced in this study are available in the following database: ENA Accession PRJEB34212 (http://www.ebi.ac.uk/ena/data/view/PRJEB34212). The WGS data produced in this study are available in the following database: ENA Accession ERP116589 (http://www.ebi.ac.uk/ena/data/view/ERP116589).

Acknowledgements

We thank Koen Schipper, Sjoerd Klarenbeek, Elizabeth Anderson, and Antoinette van Weverwijk for technical suggestions, and/or help with the experiments. We are grateful for excellent support from the NKI animal facility, animal pathology facility, transgenic facility, preclinical intervention unit, core facility molecular pathology and biobanking (CFMPB), and genomics core facility. Financial support was provided by the Oncode Institute, the Netherlands Organization for Scientific Research (NWO: Cancer Genomics Netherlands (CGCNL; grant 024001028), Cancer Systems Biology Center (CSBC; grant85300120), Netherlands Genomics Initiative (NGI) Zenith grant 93512009, VICI grant 91814643), the European Research Council (ERC Synergy project Combat Cancer, grant 319661), and a NWO National Roadmap grant for Large-Scale Research Facilities (grant 184032303).

Author contributions

Conceptualisation: SA, CL, and JJ. Experiments: SA, CL, LH, JB, KW, BS, BvG, RdK-G, APD, EvdB, TE, SM, and KB. Supervision: LFAW, DJA, MvdV, IJH, LED, and JJ. Resources: MPZ, EMS, and LED. Data analysis: SA, CL, and JJ. Writing: SA, CL, and JJ with support from DJA and LED.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1 Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47-54, doi:10.1038/nature17676 (2016).

2 Annunziato, S. *et al.* Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev.* **30**, 1470-1480, doi:10.1101/gad.279190.116 (2016).

3 Kas, S. M. *et al.* Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nat Genet* **49**, 1219-1230, doi:10.1038/ng.3905 (2017).

4 Annunziato, S. *et al.* Comparative oncogenomics identifies combinations of driver genes and drug targets in BRCA1-mutated breast cancer. *Nat Commun* **10**, 397, doi:10.1038/s41467-019-08301-2 (2019).

5 Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* **19**, 770-788, doi:10.1038/s41576-018-0059-1 (2018).

6 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).

7 Huijbers, I. J. *et al.* Rapid target gene validation in complex cancer mouse models using re-derived embryonic stem cells. *EMBO Mol Med* **6**, 212-225, doi:10.1002/emmm.201303297 (2014).

8 Liu, X. *et al.* Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1 -mutated basal-like breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12111-12116, doi:10.1073/pnas.0702969104 (2007).

9 Boelens, Mirjam C. *et al.* PTEN Loss in E-Cadherin-Deficient Mouse Mammary Epithelial Cells Rescues Apoptosis and Results in Development of Classical Invasive Lobular Carcinoma. *Cell Reports* **16**, 2087-2101, doi:10.1016/j.celrep.2016.07.059 (2016).

10 Haibe-Kains, B. *et al.* A Three-Gene Model to Robustly Identify Breast Cancer Molecular Subtypes. *JNCI: Journal of the National Cancer Institute* **104**, 311-325, doi:10.1093/jnci/djr545 (2012).

11 Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol* **36**, 888-893, doi:10.1038/nbt.4194 (2018).

12 Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292-295, doi:10.1126/science.aaw7166 (2019).

13 Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289-292, doi:10.1126/science.aav9973 (2019).

14 Kluesner, M. G. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal* **1**, 239-250, doi:10.1089/crispr.2018.0014 (2018).

15 Jiang, Y.-Z. *et al.* Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell* **35**, 428-440.e425, doi:10.1016/j.ccell.2019.02.001 (2019).

16 Hwang, G.-H. *et al.* Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* **19**, 542, doi:10.1186/s12859-018-2585-4 (2018).

17 Rottenberg, S., Pajic, M. & Jonkers, J. Studying drug resistance using genetically engineered mouse models for breast cancer. *Methods Mol Biol* **596**, 33-45, doi:10.1007/978-1-60761-416-6_3 (2010).

18 Duarte, A. A. *et al.* BRCA-deficient mouse mammary tumor organoids to study cancer-drug resistance. *Nat Methods* **15**, 134-140, doi:10.1038/nmeth.4535 (2018).

19 Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774, doi:10.1126/sciadv.aao4774 (2017).

20 Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

21 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).

22 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262, doi:10.1126/science.aas9129 (2018).

23 Huang, T. P. *et al.* Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat Biotechnol* **37**, 626-631, doi:10.1038/s41587-019-0134-y (2019).

24 Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* **37**, 276-282, doi:10.1038/s41587-018-0011-0 (2019).

25 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376, doi:10.1038/nbt.3803 (2017).

26 Jiang, W. *et al.* BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res* **28**, 855-861, doi:10.1038/s41422-018-0052-4 (2018).

27 Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat Commun* **10**, 439, doi:10.1038/s41467-018-08034-8 (2019).

28 Thuronyi, B. W. *et al.* Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol* **37**, 1070-1079, doi:10.1038/s41587-019-0193-0 (2019).

29 Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* **13**, 1036-1042, doi:10.1038/nmeth.4038 (2016).

30 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

31 Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* **34**, 631-633, doi:10.1038/nbt.3536 (2016).

32 Follenzi, A., Ailles, L. E., Bakovic, S., Geuna, M. & Naldini, L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet* **25**, 217-222, doi:10.1038/76095 (2000).

33 Derksen, P. W. B. *et al.* Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer Cell* **10**, 437-449, doi:10.1016/j.ccr.2006.09.013 (2006).

34 Krause, S., Brock, A. & Ingber, D. E. Intraductal Injection for Localized Drug Delivery to the Mouse Mammary Gland. *JoVE*, 50692, doi:10.3791/50692-v (2013).

35 Doornebal, C. W. *et al.* A Preclinical Mouse Model of Invasive Lobular Breast Cancer Metastasis. *Cancer Research* **73**, 353-363, doi:10.1158/0008-5472.CAN-11-4208 (2013).

36 Henneman, L. *et al.* Selective resistance to the PARP inhibitor olaparib in a mouse model for BRCA1-deficient metaplastic breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8409-8414, doi:10.1073/pnas.1500223112 (2015).

37 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* **42**, e168-e168, doi:10.1093/nar/gku936 (2014).

38 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

39 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

40 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

41 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

42 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

43 Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *CP in Bioinformatics* **56**, doi:10.1002/cpbi.20 (2016).

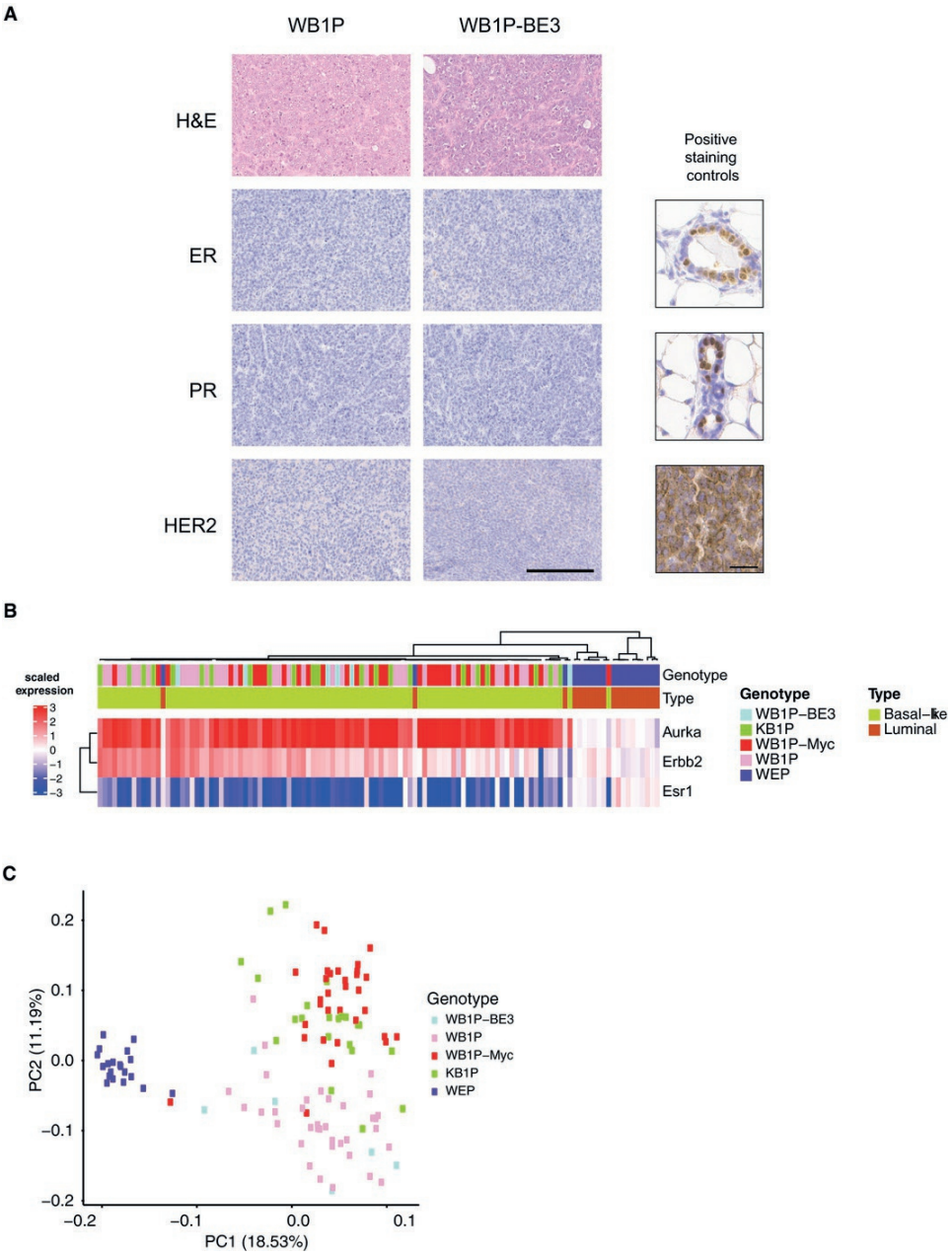
44 Raine, K. M. *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *CP in Bioinformatics* **52**, doi:10.1002/0471250953.bi1507s52 (2015).

45 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993, doi:10.1093/bioinformatics/btr509 (2011).

46 Doran, A. G. *et al.* Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. *Genome Biol* **17**, 167, doi:10.1186/s13059-016-1024-y (2016).

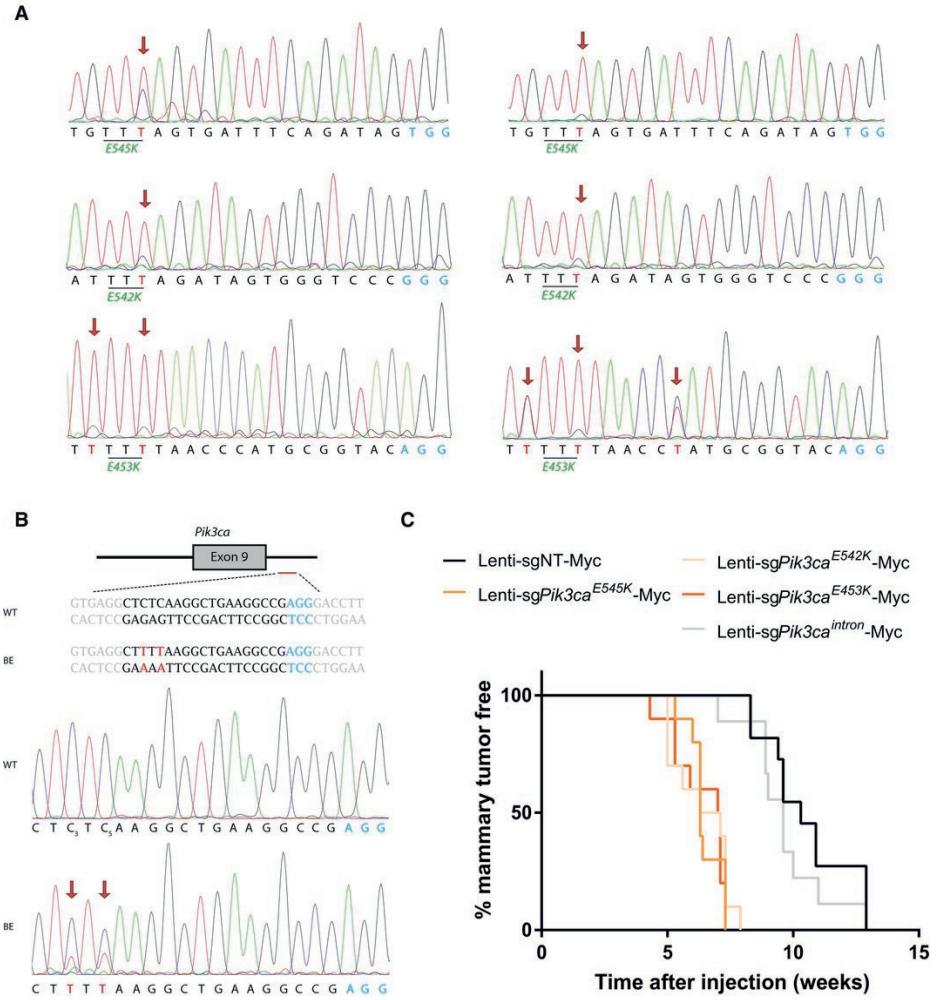
47 Menzies, A. *et al.* VAGrENT: Variation Annotation Generator. *CP in Bioinformatics* **52**, doi:10.1002/0471250953.bi1508s52 (2015).

Supplementary Data



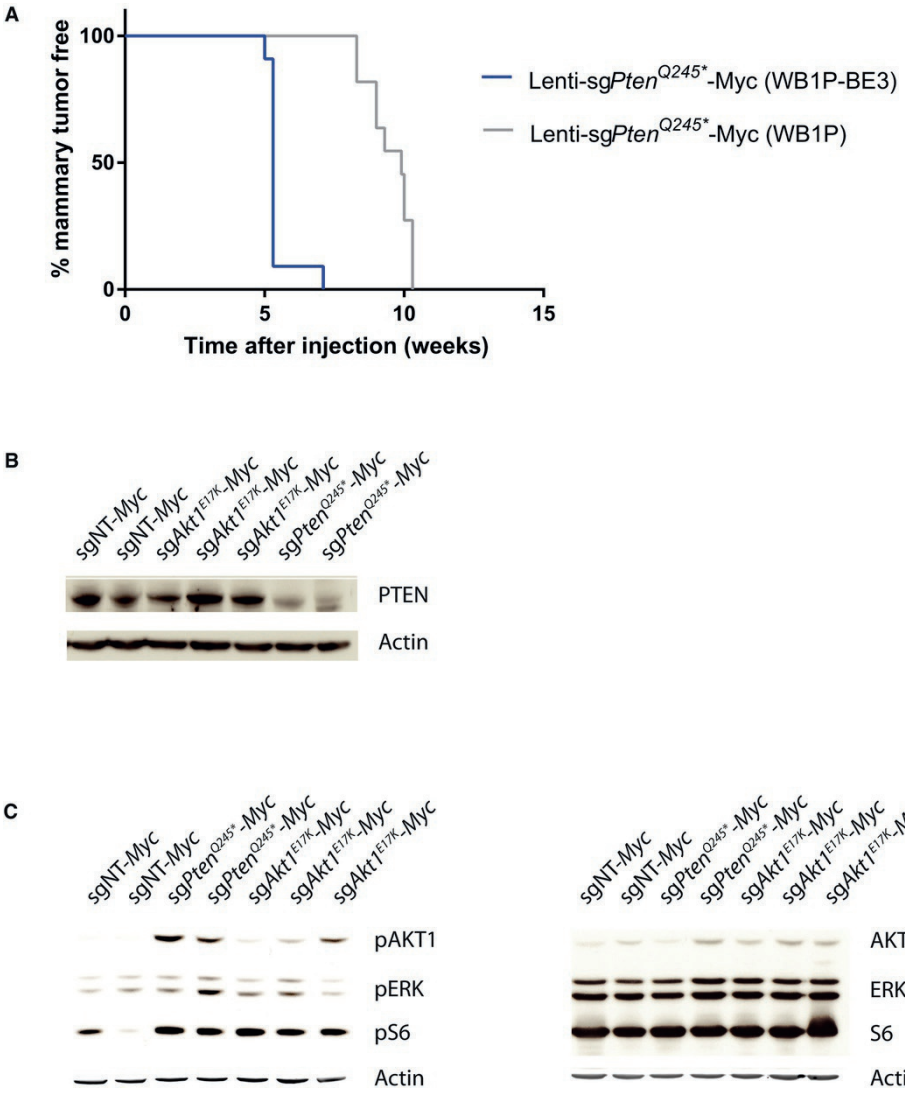
Expanded View Figure EV1: Characterisation of the WB1P-BE3 model

A) Representative hematoxylin and eosin (HE) staining and immunohistochemical detection of oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) in WB1P-BE3 tumours. Scale bar, 200 µm. Right squares display positive control stainings (scale bar, 25 µm). **B**) Unsupervised hierarchical clustering (Pearson correlation distance, average linkage) of the WB1P-BE3 tumours with tumours derived from published mouse models of luminal (*WapCre;Cdh1*$^{F/F}$*;Pten*$^{F/F}$, WEP) and basal-like (*K14Cre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$, KB1P; *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$, WB1P; *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$*;Col1a1*$^{invCAG\text{-}Myc/+}$, WB1P-Myc) breast cancer [4,8,9], using a three-gene signature that distinguishes the PAM50 subtypes [10]. **C**) PCA plot comparing WB1P-BE3 tumours to KB1P, WB1P, WB1P-Myc, and WEP tumours using global gene expression profiles.



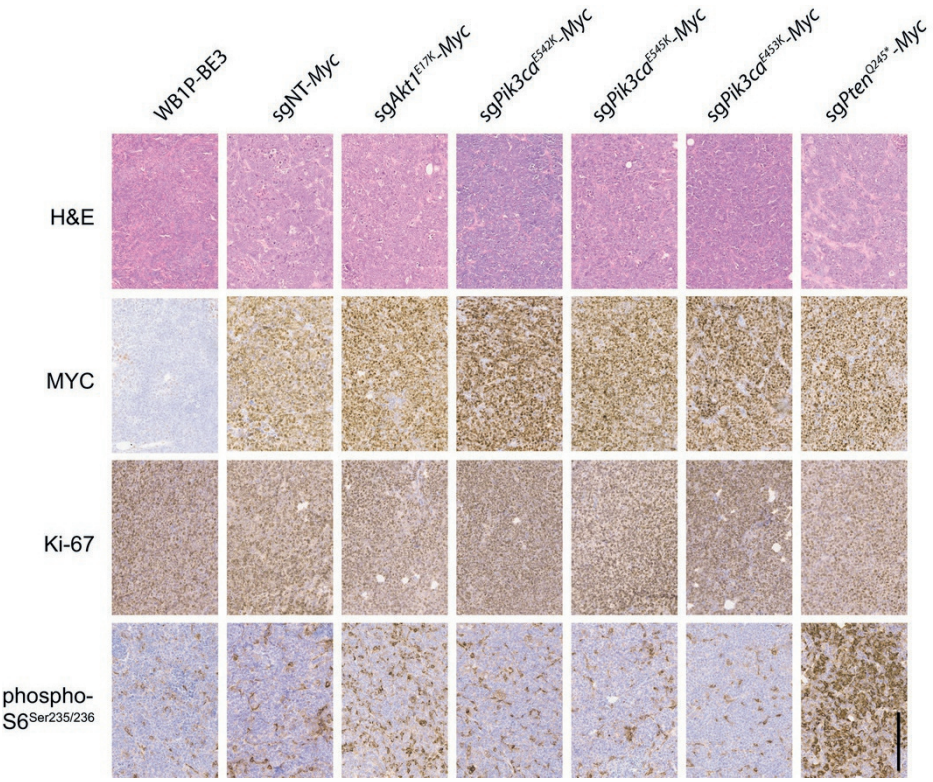
Expanded View Figure EV2: Base editing of Pik3ca

A) Sanger-sequencing chromatograms showing the target region of sg*Pik3ca*^{E542K}, sg*Pik3ca*^{E545K}, and sg*Pik3ca*^{E453K} in six independent tumours from WB1P-BE3 females injected with the corresponding Lenti-sg*Pik3ca*-Myc vectors. Arrowheads highlight cytosines of the protospacer that show base editing. For sg*Pik3ca*^{E545K} and sg*Pik3ca*^{E453K}, the same chromatograms shown in Figure 2E have been re-used here. **B**) Sanger-sequencing chromatograms showing the target region of sg*Pik3ca*^{intron} in wild-type (WT) and base edited (BE) cells. Arrowheads highlight cytosines of the protospacers that show base editing 5 days after transduction of BE3-expressing NIH3T3 cells with Lenti-sg*Pik3ca*^{intron}. **C**) Kaplan–Meier curves showing mammary tumour-specific survival for the different models. WB1P-BE3 females injected with Lenti-sg*Pik3ca*^{E542K}-Myc (n=10), Lenti-sg*Pik3ca*^{E545K}-Myc (n=10), and Lenti-sg*Pik3ca*^{E453K}-Myc (n=10) showed a reduced mammary tumour-specific survival compared to WB1P-BE3 female mice injected with Lenti-sg*Pik3ca*^{intron}-Myc (n=9) vectors (47, 44, and 49 days after injection, respectively vs. 67 days after injection, $P < 0.001$ by Mantel-Cox test). WB1P-BE3 females injected with Lenti-sgNT-Myc (n=11) showed a mammary tumour-specific survival comparable to WB1P-BE3 females injected with Lenti-sg*Pik3ca*^{intron}-Myc (not statistically significant by Mantel-Cox test).



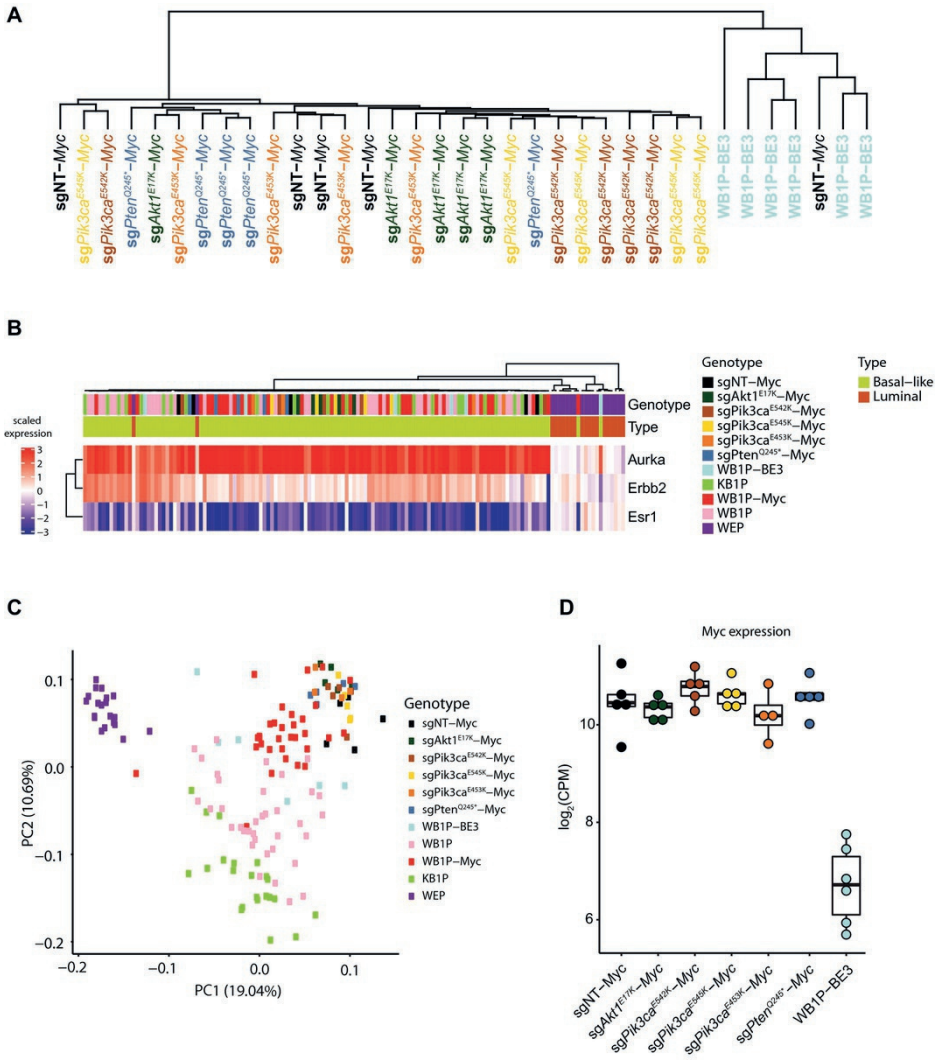
Expanded View Figure EV3: Base editing of Pten

A) Kaplan–Meier curves showing mammary tumour-specific survival for the different models. WB1P-BE3 females injected with Lenti-sg*Pten*^{Q245*}-Myc (n=11) showed a reduced mammary tumour-specific survival compared to WB1P female mice injected with Lenti-sg*Pten*^{Q245*}-Myc (n=11) vectors (37 days after injection vs. 69 days after injection, respectively $P < 0.0001$ by Mantel-Cox test). **B**) The expression of PTEN in tumours from WB1P-BE3 mice injected with Lenti-sgNT-Myc, Lenti-sg*Pten*^{Q245*}-Myc, or Lenti-sg*Akt1*^{E17K}-Myc, as visualised by immunoblotting using anti-PTEN antibody. **C**) The expression and activity of Akt1, ERK, and S6 in tumours from WB1P-BE3 mice injected with Lenti-sgNT-Myc, Lenti-sg*Pten*^{Q245*}-Myc, or Lenti-sg*Akt1*^{E17K}-Myc, as visualised by immunoblotting using anti-Akt1, anti-phospho-Akt1^{Ser473}, anti-ERK, anti-phospho-ERK (ERK1/ERK2$^{Thr202/Tyr204}$), anti-S6, and anti-phospho-S6$^{Ser235/236}$ antibodies.



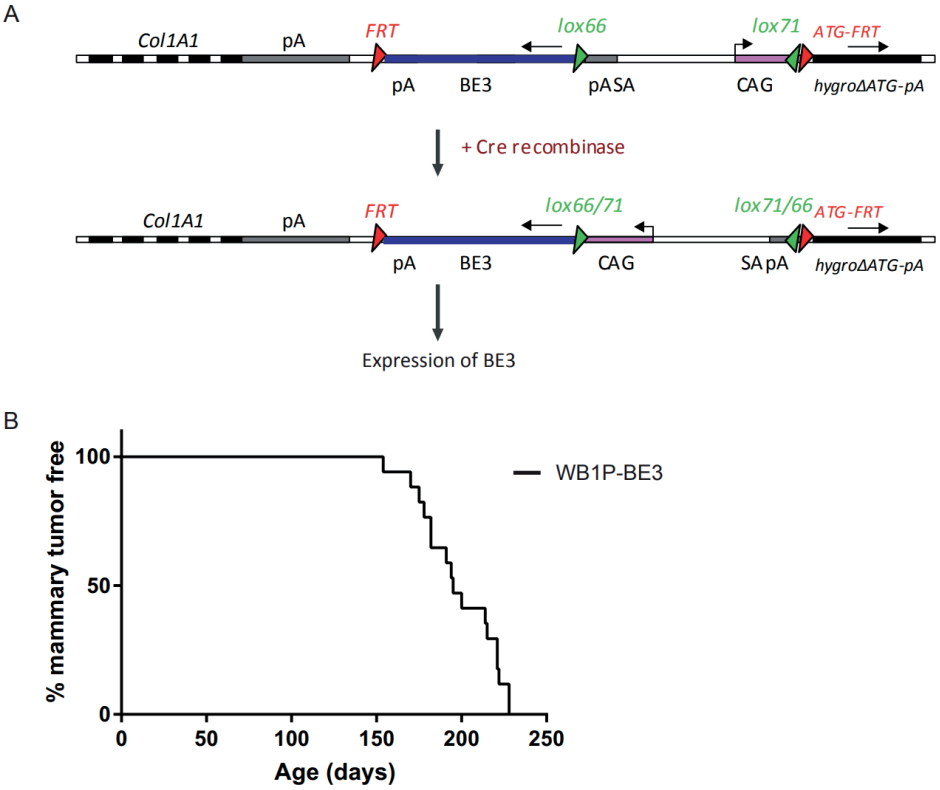
Expanded View Figure EV4: Immunohistochemical analysis of base edited tumours

Representative HE staining and immunohistochemical detection of MYC, Ki-67, and phospho-S6$^{Ser235/236}$ in spontaneous WB1P-BE3 tumours and in tumours from WB1P-BE3 females injected with the indicated Lenti-sgRNA-Myc vectors. Scale bar, 200 µm.



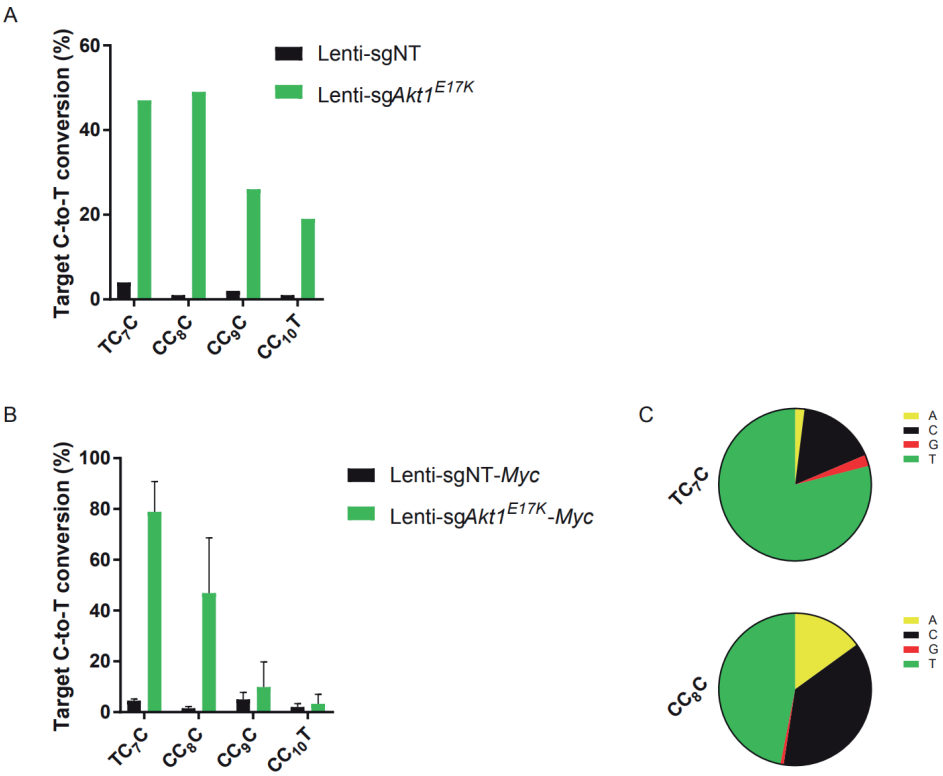
Expanded View Figure EV5: RNA-Seq analysis of base edited tumours

A) Unsupervised hierarchical clustering (Pearson correlation distance, average linkage) of the spontaneous WB1P-BE3 tumours with tumours from WB1P-BE3 mice injected with the indicated Lenti-sgRNA-Myc vectors, based on global gene expression. **B**) Unsupervised hierarchical clustering (Pearson correlation distance, average linkage) of the tumours from WB1P-BE3 mice injected with the indicated Lenti-sgRNA-Myc vectors with spontaneous WB1P-BE3 tumours and with tumours derived from published mouse models of luminal (*WapCre;Cdh1$^{F/F}$;Pten$^{F/F}$*, WEP) and basal-like (*K14Cre;Brca1$^{F/F}$;Trp53$^{F/F}$*, KB1P; *WapCre;Brca1$^{F/F}$;Trp53$^{F/F}$*, *WB1P;WapCre;Brca1$^{F/F}$;Trp53$^{F/F}$;Col1a1$^{invCAG\text{-}Myc/+}$*, WB1P-Myc) breast cancer [4,8,9], using a three-gene signature that distinguishes the PAM50 subtypes [10]. **C**) PCA plot comparing the tumours from the somatic models with spontaneous WB1P-BE3 tumours and to KB1P, WB1P, WB1P-Myc, and WEP tumours, using global gene expression profiles. **D**) Myc expression levels in spontaneous WB1P-BE3 tumours and in tumours from WB1P-BE3 mice injected with the indicated Lenti-sgRNA-Myc vectors (n=4-6). Boxes extend from the third (Q3) to the first (Q1) quartile (interquartile range, IQR), with the line at the median; whiskers extend to Q3 + 1.5× IQR and to Q1 - 1.5× IQR.



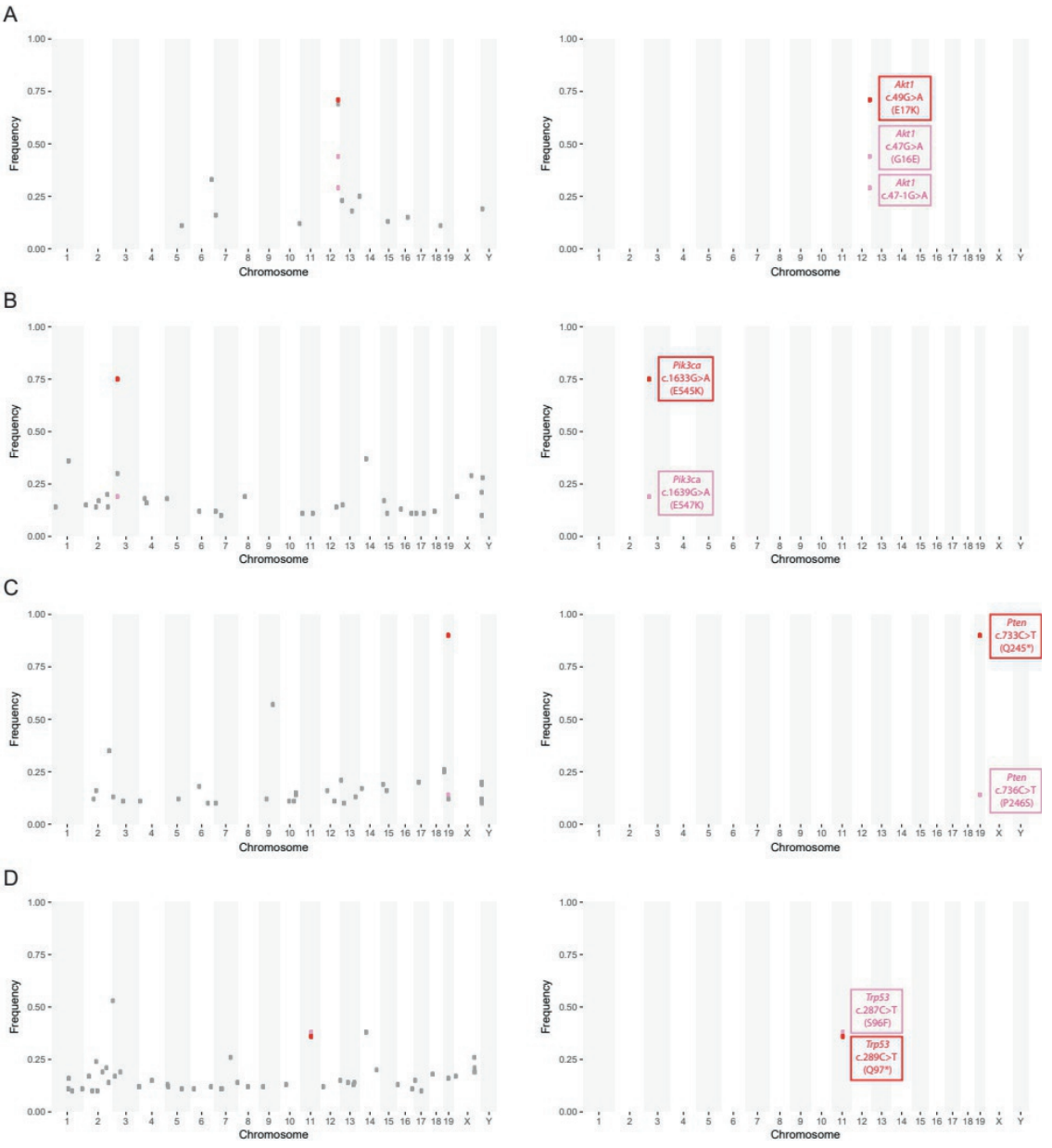
Appendix Figure S1: Characterisation of the WB1P-BE3 model

A) Depiction of the Cre-conditional invCAG-BE3 allele integrated into the Col1a1 locus of embryonic stem cells (ESCs) derived from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$ mice. WapCre-mediated recombination allows mammary-specific inversion of the CAG promoter, resulting in expression of BE3. **B**) Kaplan-Meier curve showing mammary tumour-specific survival for *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/F}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ (WB1P-BE3) female mice (n=17).



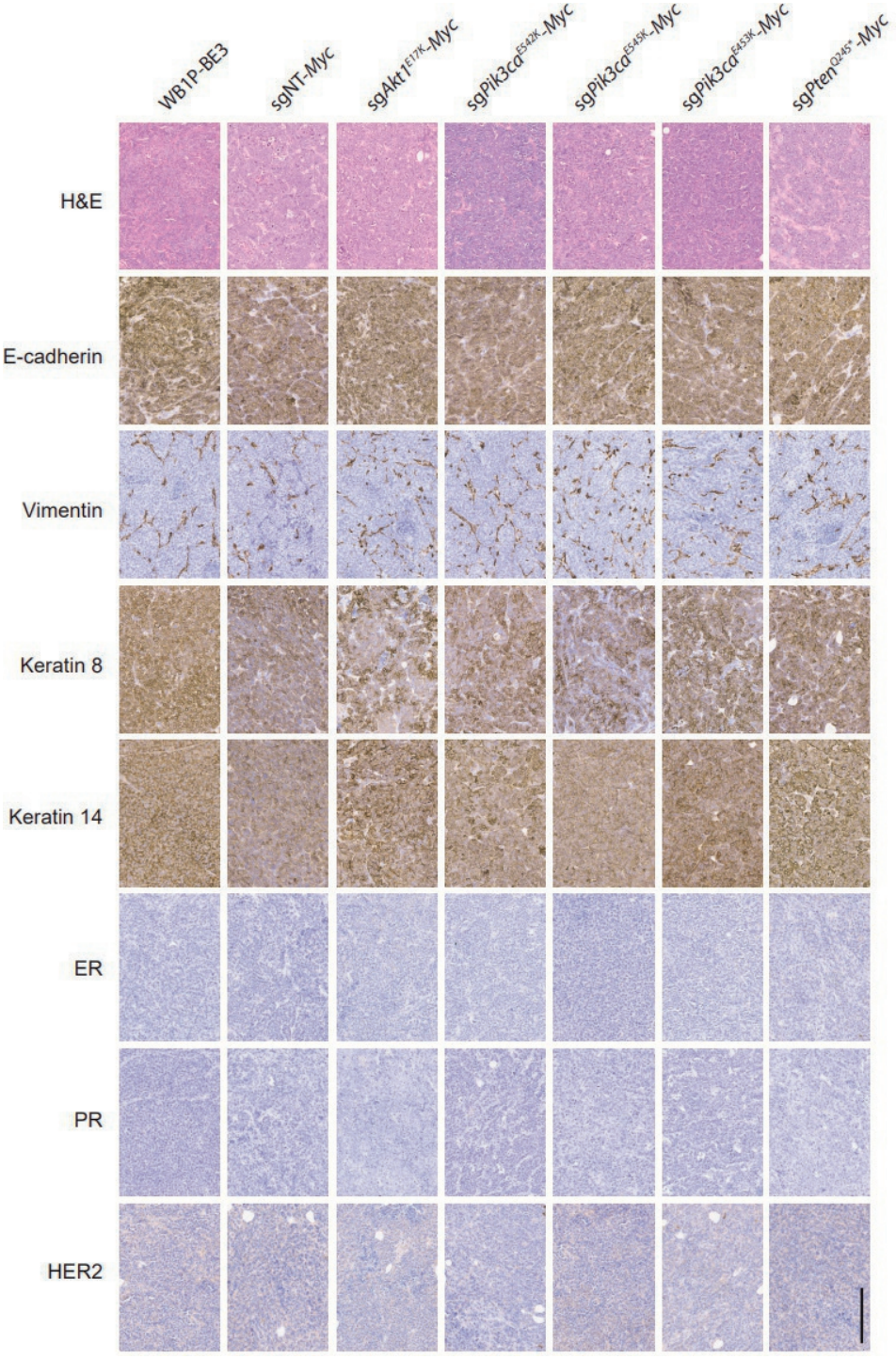
Appendix Figure S2: Base editing of Akt1

A) EditR [14] was used to calculate the frequency (%) of C-to-T conversion at the indicated target cytosines of the protospacer in BE3-expressing NIH3T3 cells 5 days after transduction with the indicated sgRNA vectors. **B**) EditR was used to calculate the average frequency (%) of C-to-T conversion at the indicated target cytosines of the protospacer in tumours from WB1P-BE3 females injected with Lenti-sgNT-Myc or Lenti-sg*Akt1*^{E17K}-Myc. **C**) Pie chart showing, in tumours from WB1P-BE3 animals injected with Lenti-sg*Akt1*^{E17K}-Myc, the average product purity at the indicated target cytosines of the protospacer. Note that neither of the possible base substitutions at C8 can change the amino acid sequence of *AKT1* (for its wobble codon position).


A
Frequency
Chromosome
Akt1 c.49G>A (E17K)
Akt1 c.47G>A (G16E)
Akt1 c.47-1G>A
B
Pik3ca c.1633G>A (E545K)
Pik3ca c.1639G>A (E547K)
C
Pten c.733C>T (Q245*)
Pten c.736C>T (P246S)
D
Trp53 c.287C>T (S96F)
Trp53 c.289C>T (Q97*)

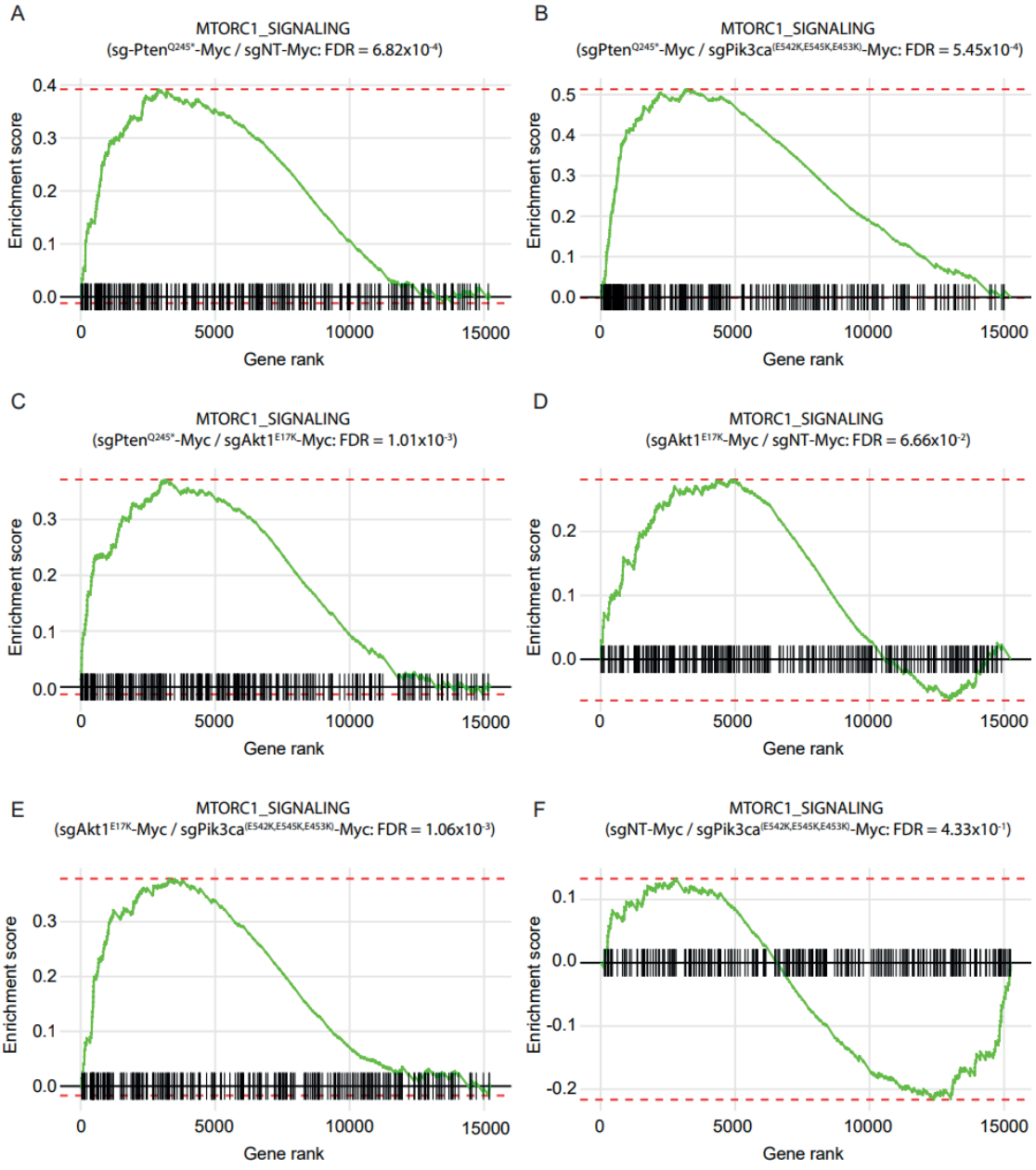
Appendix Figure S3: Genome-wide off-target analysis of base edited cells

Upon whole-genome sequencing (WGS) of DNA isolated from NIH3T3 cells transduced with the CBE and either Lenti-sg*Akt1*^{E17K}-Myc (A), Lenti-sg*Pik3ca*^{E545K}-Myc (B), Lenti-sg*Pten*^{Q245*}-Myc (C) or Lenti-sg*Trp53*^{Q97*}-Myc (D), on-target base edits could be readily detected at high allele frequencies (y-axis). In addition to this, a limited number of additional SNVs was identified (left panels), but none of these off-target edits generated missense or nonsense mutations or altered essential splice sites (right panels). Gray, red and pink dots indicate off-target SNVs, on-target edits and bystander mutations at nearby cytosines, respectively.


WB1P-BE3
sgNT-Myc
sgAkt1E17K-Myc
sgPik3caE542K-Myc
sgPik3caE545K-Myc
sgPik3caH453K-Myc
sgPtenQ245*-Myc
H&E
E-cadherin
Vimentin
Keratin 8
Keratin 14
ER
PR
HER2

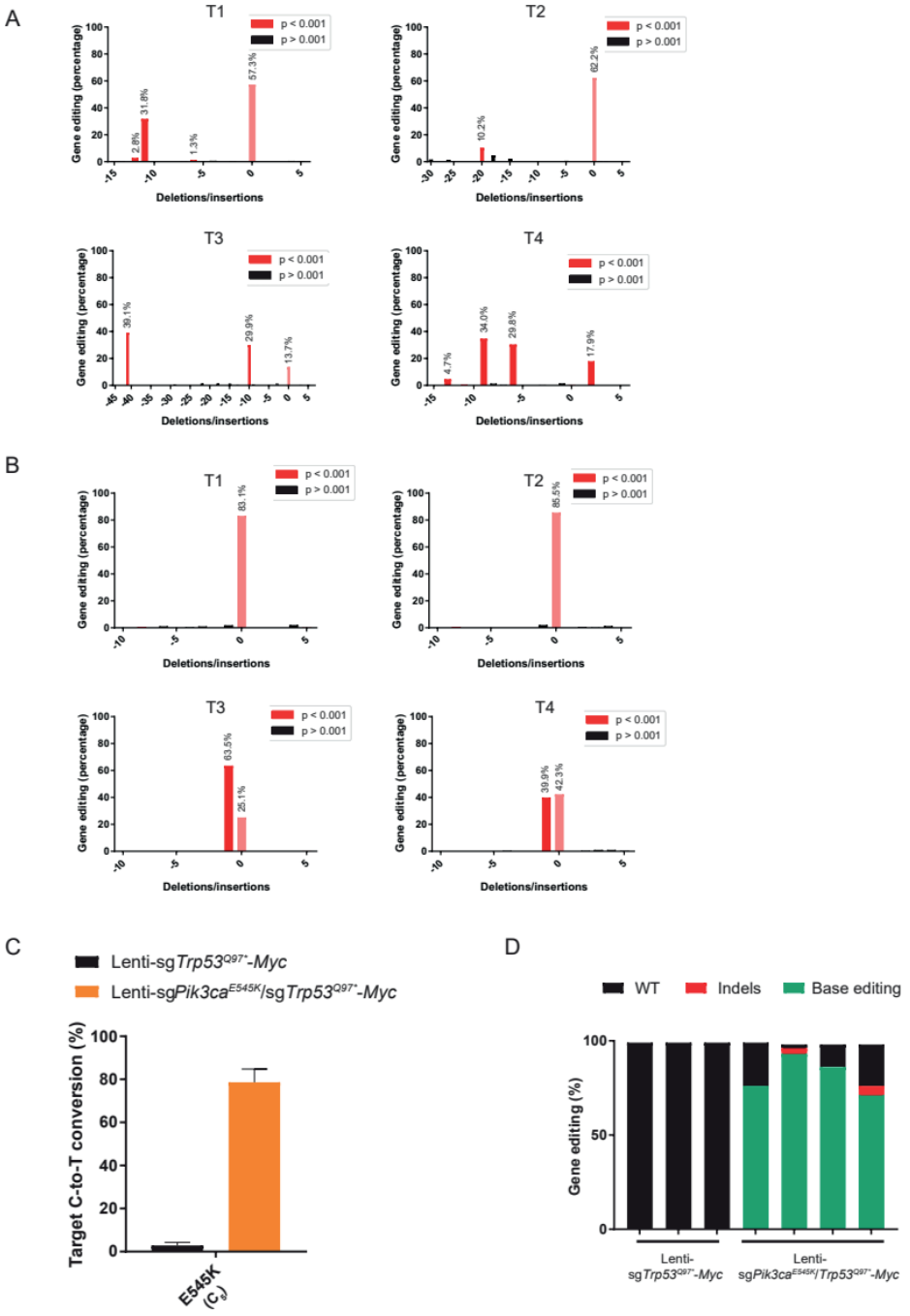
Appendix Figure S4: Immunohistochemical analysis of base edited tumours

Representative HE staining (re-used from Figure EV4) and immunohistochemical detection of E-cadherin, vimentin, keratin 8, keratin 14, ER, PR and HER2 in spontaneous WB1P-BE3 tumours and in tumours from WB1P-BE3 females injected with the indicated Lenti-sgRNA-Myc vectors. Bar, 200 µM. For ER, PR and HER2 positive staining controls, refer to Figure EV1A.


A
MTORC1_SIGNALING
(sg-Pten^{Q245*}-Myc / sgNT-Myc: FDR = 6.82x10^{-4})
B
MTORC1_SIGNALING
(sgPten^{Q245*}-Myc / sgPik3ca$^{(E542K,E545K,E453K)}$-Myc: FDR = 5.45x10^{-4})
C
MTORC1_SIGNALING
(sgPten^{Q245*}-Myc / sgAkt1^{E17K}-Myc: FDR = 1.01x10^{-3})
D
MTORC1_SIGNALING
(sgAkt1^{E17K}-Myc / sgNT-Myc: FDR = 6.66x10^{-2})
E
MTORC1_SIGNALING
(sgAkt1^{E17K}-Myc / sgPik3ca$^{(E542K,E545K,E453K)}$-Myc: FDR = 1.06x10^{-3})
F
MTORC1_SIGNALING
(sgNT-Myc / sgPik3ca$^{(E542K,E545K,E453K)}$-Myc: FDR = 4.33x10^{-1})
Enrichment score
Gene rank
0
5000
10000
15000

Appendix Figure S5: Gene set enrichment plot for "MTORC1_SIGNALING" from the MSigDB Hallmark gene set

X-axis indicates the sorted genes according to the moderated T-statistics in the comparison of Lenti-sg*Pten*^{Q245*}-Myc vs Lenti-sgNT-Myc (A), Lenti-sg*Pten*^{Q245*}-Myc vs Lenti-sg*Pik3ca*$^{(E542K,E545K,E453K)}$-Myc (B), Lenti-sg*Pten*^{Q245*}-Myc vs Lenti-sg*Akt1*^{E17K}-Myc (C), Lenti-sg*Akt1*^{E17K}-Myc vs Lenti-sgNT-Myc (D), Lenti-sg*Akt1*^{E17K}-Myc vs Lenti-sg*Pik3ca*$^{(E542K,E545K,E453K)}$-Myc (E), Lenti-sgNT-Myc vs Lenti-sg*Pik3ca*$^{(E542K,E545K,E453K)}$-Myc (F) using limma-voom. Y-axis indicates enrichment score and red dotted line indicates the maximum and minimum enrichment scores by the gene set.


A
T1
T2
T3
T4
p < 0.001
p > 0.001
Gene editing (percentage)
Deletions/insertions
2.8%
31.8%
1.3%
57.3%
10.2%
62.2%
39.1%
29.9%
13.7%
4.7%
34.0%
29.8%
17.9%
B
83.1%
85.5%
63.5%
25.1%
39.9%
42.3%
C
Lenti-sgTrp53Q97*-Myc
Lenti-sgPik3caE545K/sgTrp53Q97*-Myc
Target C-to-T conversion (%)
E545K (C9)
D
WT
Indels
Base editing
Gene editing (%)
Lenti-sgTrp53Q97*-Myc
Lenti-sgPik3caE545K/Trp53Q97*-Myc

Appendix Figure S6: Analysis of unintended by-product indel formation

A) TIDE analysis showing the spectrum of insertions/deletions (indels) of the targeted *Pten* alleles in four independent tumours from WB1P-BE3 mice injected with Lenti-sg*Pten*^{Q245*}-Myc. **B**) TIDE analysis showing the spectrum of indels of the targeted *Trp53* alleles in four independent tumours from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ mice injected with Lenti-sg*Trp53*^{Q97*}-Myc or Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc. **C**) EditR was used to calculate the average frequency (%) of C-to-T conversion at C5 of the protospacer targeted by sg*Pik3ca*^{E545K} in tumours from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ females injected with Lenti-sg*Trp53*^{Q97*}-Myc or Lenti-sg^{Pik3caE545K}/sg*Trp53*^{Q97*}-Myc. **D**) BE Analyzer [16] was used to assess from next-generation sequencing data the fraction of wild-type *Pik3ca* alleles, base edited alleles or alleles with indels in tumours from *WapCre;Brca1*$^{F/F}$*;Trp53*$^{F/+}$*;Col1a1*$^{invCAG\text{-}BE3/+}$ animals injected with Lenti-sg*Trp53*^{Q97*}-Myc or Lenti-sg*Pik3ca*^{E545K}/sg*Trp53*^{Q97*}-Myc.