Early detection of pre-malignant lesions in a KRAS^{G12D}-driven mouse lung cancer model by monitoring circulating-free DNA

Callum P Rakhit[1§], Ricky M Trigg[2§*], John Le Quesne[1,2], Michael Kelly[3], Jacqueline A Shaw[2‡], Catrin Pritchard[2‡], L. Miguel Martins[1‡]

[1] MRC Toxicology Unit, University of Cambridge, Lancaster Road, Leicester LE1 9HN, UK

[2] Leicester Cancer Research Centre, University of Leicester, Leicester LE2 7LX

[3]Core Biotechnology Services, University of Leicester, Leicester LE1 7RH

*Present address: Division of Cellular and Molecular Pathology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ

[‡]Authors for correspondence: L. Miguel Martins, martins.lmiguel@gmail.com, Catrin Pritchard, cap8@le.ac.uk and Jacqueline Shaw, js39@le.ac.uk

[§]These authors contributed equally to this work

Running title: cfDNA detection in *Kras^{G12D}* mouse lung model

Key words: KRAS^{G12D}, mouse model, circulating-free DNA (cfDNA), early detection, lung adenocarcinoma

Summary statement (15-30 words): Developing a liquid biopsy to track early lung cancer in a mouse model driven by *KRAS* mutation

ABSTRACT

Lung cancer is the leading cause of cancer-related death. Two-thirds of cases are diagnosed at an advanced stage that is refractory to curative treatment. Therefore, strategies for the early detection of lung cancer are urgently sought. Total circulating-free DNA (cfDNA) and tumour-derived circulating tumour DNA (ctDNA) are emerging as important biomarkers within a 'liquid biopsy' for monitoring human disease progression and response to therapy. Due to the late clinical diagnosis of lung adenocarcinoma, the potential for cfDNA and ctDNA as early detection biomarkers remains unexplored. Here, using a Cre-regulated genetically engineered mouse model of lung adenocarcinoma development, driven by $Kras^{G12D}$ (the $Kras^{LSL\text{-}G12D}$ mouse), we serially track the release of cfDNA/ctDNA and compare this to tumour burden as determined by micro-computed tomography (micro-CT). To monitor ctDNA, a droplet digital PCR assay was developed to permit discrimination of the $Kras^{Lox\text{-}G12D}$ allele from the $Kras^{LSL\text{-}G12D}$ and $Kras^{WT}$ alleles. We show that micro-CT correlates with endpoint histology and is able to detect pre-malignant tumours with a combined volume larger than $7mm^3$. Changes in cfDNA/ctDNA levels correlate with micro-CT measurements in longitudinal sampling and are able to monitor the emergence of lesions before the adenoma-adenocarcinoma transition. This work has implications for early detection of human lung adenocarcinoma using ctDNA/cfDNA profiling.

INTRODUCTION

Circulating cell free DNA (cfDNA) was first identified in the human bloodstream during the first half of the 20th century (Mandel and Metais, 1948), though only in the late 1980s was cfDNA, isolated from the plasma of cancer patients, shown to be partially derived from tumours (Stroun et al., 1987), representing the so-called circulating tumour DNA or ctDNA fraction. ctDNA is a reliable biomarker for identifying oncogenic changes within the body (Schwarzenbach et al., 2011) and has implications for both the early detection and monitoring of cancer. Changes in the molecular profile of ctDNA can be used to detect early-stage cancer lesions (Cheng et al., 2017), classify the molecular profiles of existing tumours (Burrell et al., 2013), identify the emergence of resistance (Diaz et al., 2012), and track the evolution of cancer genomes in response to targeted drug therapies (Abbosh et al., 2017). In this regard, ctDNA analysis is transforming the monitoring of cancer after diagnosis and has been established as a prognostic factor for lung cancer patients (Cabral et al., 2010; Goebel et al., 2005).

Monitoring a patient's ctDNA profile is less invasive than tissue biopsy, with the reduced stress and cost allowing for more frequent sampling. Repeat sampling allows for more targeted, personalised therapy in the face of tumour evolution. ctDNA may also have the potential to enable the discovery of newly emergent cancers, undetectable by imaging or other diagnostic procedures (Abbosh et al., 2017; Chaudhuri et al., 2017; Garcia-Murillas et al., 2015). Accordingly, ctDNA profiling approaches are being developed to improve the monitoring and characterisation of residual disease in lung cancer, which may help improve outcomes in the adjuvant disease setting (Abbosh et al., 2017). However, the genetic profiling of ctDNA in individuals with early disease lesions has been limited, due to

inherent challenges in building a cohort of study patients with identifiable pre-malignant lesions.

Gain-of-function mutations in *KRAS* are present in ~25% of human lung adenocarcinomas and are truncal events, acquired early in disease development (Abbosh et al., 2017; Rakhit et al., 2017). Oncogenic mutants of *KRAS*, such as the prevalent ^{G12D}*KRAS* mutation, have transforming activity and are thought to be founder mutations since they can initiate and drive tumour progression in mouse models (Guerra et al., 2003; Jackson et al., 2001; Sutherland et al., 2014). The autochthonous $Kras^{LSL\text{-}G12D}$ conditional mouse knock-in model, which allows for endogenous expression of ^{G12D}Kras following Cre induction, has been used extensively to study the mechanisms underpinning early disease initiation and maintenance (Jackson et al., 2005; Jackson et al., 2001; Sheridan and Downward, 2015; Sutherland et al., 2014). Evidence shows that the model recapitulates early stage lung adenocarcinoma development, through the formation of atypical adenomatous hyperplasia (AAH), epithelial hyperplasia of the bronchioles and papillary adenomas (Jackson et al., 2001; Nikitin et al., 2004). Early stage lung adenocarcinomas are occasionally seen at late stages, following prolonged ^{G12D}Kras expression, and can be accelerated by a combined p53 mutation (Jackson et al., 2005).

Here, we use the $Kras^{LSL\text{-}G12D}$ mouse model to explore the utility of both total cfDNA levels and ctDNA as an early-stage biomarker. We show that cfDNA/ctDNA is detectable in mice bearing pre-malignant lung lesions, prior to the adenocarcinoma transition, and we correlate the liquid biopsy data with the emergence of pre-malignant lung lesions as detected by longitudinal micro-CT.

RESULTS

Comparing tumour size and number in *Kras*$^{+/Lox\text{-}G12D}$ mice using micro-CT and histology

To induce expression of KRas^{G12D} in the mouse lung, adenoviral-Cre vectors are delivered to *Kras*$^{+/LSL\text{-}G12D}$ mice using intranasal delivery (Fig 1A) and the most commonly used viruses used for this purpose are either Ad5-CMV-Cre or Ad5-SPC-Cre (Jackson et al 2001; Sutherland et al 2014). With Ad5-CMV-Cre, the Cre recombinase is expressed from the ubiquitous CMV promoter and generates a range of lung pathologies including AAH/adenoma and bronchial hyperplasia (BH) (Jackson et al., 2001). Ad5-CMV-Cre is also known to induce recombination of the *Kras*$^{LSL\text{-}G12D}$ allele in lung resident myeloid cells and features of the Langerhans cell histiocytosis (LCH) phenotype (Kamata et al., 2017). In the case of Ad5-mSPC-Cre, the Cre recombinase is expressed from the mouse surfactant protein C (SPC) promoter allowing more restricted expression of ^{G12D}Kras to alveolar type II (ATII) cells and inducing the development of AAH/adenoma (Sutherland et al., 2014). With both models, overt adenocarcinomas are detectable at a low frequency over prolonged periods (Jackson et al., 2001; Sutherland et al., 2014). A previous study has shown a good correlation between micro-CT and histology using the *Kras*$^{+/LSL\text{-}G12D}$ model infected with Ad5-CMV-Cre. Therefore, we focussed on extending this analysis to *Kras*$^{+/LSL\text{-}G12D}$ mice infected with Ad5-mSPC-Cre.

We examined the dynamic range for tumour detection in *Kras*$^{+/LSL\text{-}G12D}$ mice 40 weeks post infection (p.i.) with Ad5-mSPC-Cre by comparing the tumour burden at endpoint using histology and micro-CT (Fig 1A). The lesions detected at this endpoint were heterogeneous in size (Fig 1A, iii). We found that the number of tumours observed by histology was significantly higher ($p < 0.01$) than observed

when micro-CT was used (Fig 1B). A further analysis, where tumours were grouped according to volume using data binning revealed that the sensitivity of micro-CT for tumour detection was lost when tumour volume fell below 0.5 mm^3, with micro-CT showing a significantly smaller ($p < 0.05$) tumour number (Fig 1C, D).

To ensure that micro-CT provides an accurate indication of individual tumour volume for tumours with volumes above 0.5 mm^3, we compared tumours assessed by micro-CT to area quantitation determined by serial sectioning followed by H&E staining. The two methods showed a linear correlation (Fig 1E), indicating that both methods were concordant for tumour detection above a 0.5-mm^3 threshold. Overall, our data show that micro-CT provides an accurate value for tumour volumes above a lower limit of 0.5 mm^3, consistent with data from Kirsch and colleagues (Kirsch et al., 2010).

Analysis of tumour histology types

To gain an assessment of the spectrum of lung tumour histology types, we undertook histopathological evaluation of H&E-stained lung sections from *Kras*$^{+/LSL\text{-}G12D}$ mice at 20 weeks p.i. with Ad5-CMV-Cre and at 40 weeks p.i. induction with Ad5-mSPC-Cre, using the recommended criteria of mouse lung tumour classification (Nikitin et al., 2004). Figure 2A provides representative images of H&E-stained sections of AAH, adenoma and bronchial lesions in Ad5-CMV-Cre infected mice. As previously reported (Jackson et al., 2001; Sutherland et al., 2014), our analysis showed the presence of AAH and adenomas in both sets of mice. Adenomas were mostly of simple papillary pattern but rare small foci of solid growth were identified. Bronchial lesions were evident in Ad5-CMV-Cre infected mice, as expected, but were largely absent from Ad5-SPC-Cre (Fig 2B). Lesions of each histological type

were counted (Fig 2B) and this showed more AAHs/adenomas in the Ad5-CMV-Cre infected mice compared to Ad5-mSPC-Cre infected mice. However, there was a high degree of inter-mouse variability, regardless of the type of virus used for reasons that are likely related to some variability in the amount of virus inhaled by individual mice. Interestingly, no overt adenocarcinomas were observed in either the Ad5-CMV-Cre or Ad5-mSPC-Cre infected *Kras*$^{+/LSL\text{-}G12D}$ mice at the time points analysed.

In vivo monitoring of tumour burden by measuring levels of total cfDNA

We next assessed whether non-invasive monitoring of lung tumour development could be achieved by profiling levels of total cfDNA. We infected *Kras*$^{+/LSL\text{-}G12D}$ mice with viruses carrying either Ad5-CMV-Cre or Ad5-CMV-βgal (control) for up to 20 weeks (Fig 3A) or with Ad5-mSPC-Cre for up to 40 weeks (Fig 3B) by intranasal delivery and monitored tumour development every 2 weeks in the case of Ad5-CMV-Cre/Ad5-CMV-βgal and every 5 to 10 weeks in the case of Ad5-mSPC-Cre by micro-CT scanning (Fig 3, right panels). We collected blood samples for cfDNA analysis, in parallel with the micro-CT imaging.

We observed a more rapid increase in tumour volume with Ad5-CMV-Cre infection (Fig 3A) than with Ad5-mSPC-Cre (Fig 3B) infection, which was consistent with previously reported data (Jackson et al., 2001; Sutherland et al., 2014) and the data shown in Figure 2B. In Ad5-CMV-Cre-infected mice, tumours larger than 0.5 mm^3 were first detectable by micro-CT at approximately 9 weeks post-infection (p.i.) (Fig 3A), whereas in Ad5-mSPC-Cre infected mice, tumours above this threshold were detectable beginning at 35 weeks p.i. (Fig 3B).

To assess total cfDNA levels in blood, a qPCR assay targeting a 113 bp single-copy locus within the *Gapdh* genomic region was used. This assay has a dynamic range of at least 500-fold and can detect the equivalent of one copy (3.3 pg; one haploid genome equivalent) of *Gapdh* (Trigg, 2017). As expected, given the lower tumour burden of the Ad5-mSPC-Cre mice, the cfDNA levels induced by Ad5-mSPC-Cre were lower than those induced by Ad5-CMV-Cre (Fig 3A and 3B, left panels). However, for tumours of equivalent volumes, cfDNA levels in Ad5-CMV-Cre- and Ad5-mSPC-Cre-infected mice were broadly similar, suggesting that $Kras^{G12D}$ expression in lung resident myeloid cells or BH induced by Ad5-CMV-Cre had little effect on overall cfDNA release.

A significant increase in cfDNA levels was first detected at 12 weeks p.i. in the case of Ad5-CMV-Cre (Fig 3A) in comparison to samples at baseline. However, in the case of Ad5-mSPC-Cre, a significant increase in cfDNA levels was first detectable at 20 weeks p.i. (Fig 3B) but this was not maintained until after 35 weeks p.i. For Ad5-CMV-βgal, cfDNA levels remained consistently low throughout the time course (Fig 3A). These data show that overall cfDNA levels were increased in mice bearing lesions that were representative of early pre-malignant lung lesions, without transition to adenocarcinoma, compared to healthy controls. For both adenoviral systems, cfDNA levels, as measured by qPCR analysis, did not reproducibly improve the threshold for tumour burden compared to micro-CT scanning, although the data for the 20-week time point in Ad5-mSPC-Cre mice potentially suggests a more sensitive cfDNA assay method may facilitate this.

Development of a PCR assay to monitor ctDNA through *Kras$^{Lox\text{-}G12D}$*

Although cfDNA levels can give an indication of tumour burden, their assessment is not able to distinguish tumour-derived ctDNA from that derived from apoptosis of other healthy cells. Therefore, we developed an assay for measuring ctDNA through the detection of the tumour-derived *Kras$^{Lox\text{-}G12D}$* allele in blood. Initial studies using endpoint PCR approaches detected the presence of this allele in the cfDNA of Ad5-CMV-Cre-infected *Kras$^{+/LSL\text{-}G12D}$* mice at 12 weeks p.i. (Trigg, 2017), encouraging the development of a quantitative PCR assay.

An assay that specifically detects the *Kras$^{Lox\text{-}G12D}$* allele was required, but this proved challenging due to similarities with the *Kras$^{LSL\text{-}G12D}$* and *Kras^{WT}* alleles. A unique 34-bp region that contains the single LoxP sequence and has 13-bp palindromic motifs flanking a central 8-bp spacer sequence was identified (Fig S1A). This sequence was determined to have a propensity to form a stem-loop structure (Fig S1B) that occludes access to primers and DNA polymerase (Huang et al., 2007). Moreover, this LoxP sequence is flanked by palindromic *Sal*1 restriction sites, which contribute a further 6 bp to the stem-loop structure; this 19 bp structure (Fig S1C) has a predicted melting temperature (T_m) of 65.5°C by *in silico* prediction using mFold (Zuker, 2003). To overcome the inhibitory effect of this stem-loop on PCR amplification, several strategies were attempted.

The first PCR strategy involved targeting the stem-loop with dual-labelled TaqMan probes incorporating a 3' minor groove binder moiety (Fig S2, MGB-1, MGB-2 or MGB-3). However, this failed to permit PCR amplification, even when destabilising agents such as betaine, urea and DMSO were included (data not shown). A second strategy was taken in which a hydrolysis probe was designed to include 'locked nucleic acid' (LNA) bases, a chemical modification that increases

both the T_m and binding specificity of the probe while also providing strand-invasion properties. This LNA probe, LNA-1 (Fig. S2), permitted qPCR amplification of the *Kras*$^{Lox\text{-}G12D}$ allele, but the efficiency was poor (~85%; Fig S3A and B). To attempt to improve the amplification efficiency, we next adapted the LNA-1 probe to a droplet digital PCR (ddPCR) assay. While this approach provided some discrimination between positive and negative droplets, there was an additional clustering of droplets above the main negative cluster, above 5000 fluorescence units (Fig S4), indicating the presence of artefacts caused by autofluorescence.

A second LNA probe, LNA-2 (Fig S2), was next designed to destabilise both stem-loops simultaneously, using a 'double destabilisation' strategy (Esposito et al., 2003). The combination of probe LNA-2 with the 133-bp amplicon primers provided good discrimination between positive and negative droplets with *Kras*$^{+/Lox\text{-}G12D}$ genomic DNA as a template (Fig 4A-D). To determine the sensitivity of the LNA-2 assay, *Kras*$^{+/Lox\text{-}G12D}$ genomic DNA was spiked into a background of *Kras*$^{+/LSL\text{-}G12D}$ genomic DNA and serially diluted two-fold over eleven points (Table S1). To ensure that the concentration of *Kras*$^{LSL\text{-}G12D}$ in each dilution was constant, the LNA-2 assay was duplexed with the *Gapdh* assay, and the copy number of *Kras*$^{Lox\text{-}G12D}$ was calculated by halving that of *Gapdh*. A good correlation between the actual and theoretical copy number was observed down to the tenth dilution, which corresponded to a copy number of 0.75 and allele frequency of 0.1% (Table S1; Fig 3E). The dropout at the eleventh dilution was consistent with only a 38% chance of a positive droplet being present, thus confirming the suitability of the LNA-2 assay for the sensitive detection of *Kras*$^{Lox\text{-}G12D}$ in plasma.

Cre-mediated recombination introduces mutations in the *Kras$^{Lox\text{-}G12D}$* allele

Cre recombinase is known to have DNA damaging activity (Loonstra et al., 2001), but its potential for introducing mutations at target recombination sites, within the *Kras$^{LSL\text{-}G12D}$* allele, has not been investigated previously. To investigate whether possible sequence variation as a result of Cre expression may impact PCR sensitivity, DNA was extracted from a FFPE section of *Kras$^{+/Lox\text{-}G12D}$* mouse lung tissue containing multiple adenomas as determined by histological analysis (Fig S5A). A 180-bp region surrounding the recombined LoxP sequence was PCR-amplified and cloned, and 40 individual clones were sequenced. Sequence alignment revealed point mutations within 12 of the 40 clones, clustered within and around the *Sal*1 restriction sites (Fig S5B, C). To confirm that these mutations were not induced by errors from the DNA polymerase during PCR amplification, a 140-bp WT region of the *Kras^{WT} allele* co-amplified by PCR in the same reaction was also cloned and sequenced, but no mutations were identified (data not shown). Thus, Cre recombination of LoxP sequences introduces point mutations within target sequences that may have an impact on the efficiency of PCR amplification.

Monitoring ctDNA by measuring *Kras$^{Lox\text{-}G12D}$* levels in plasma

We next applied the ddPCR assay developed above to monitor levels of the *Kras$^{Lox\text{-}G12D}$* allele in the plasma ctDNA fraction of *Kras$^{+/LSL\text{-}G12D}$* mice infected with Ad5-mSPC-Cre over a time course of 0 to 40 weeks. We chose to perform the analysis in Ad5-mSPC-Cre-induced mice only, since Ad5-CMV-Cre is known to induce recombination of the *Kras$^{LSL\text{-}G12D}$* allele in non-tumour cells (Kamata et al., 2017). We observed a significant increase in the number of copies of the *Kras$^{Lox\text{-}G12D}$* allele at 40 weeks p.i compared to samples at baseline, whereas an increase in

tumour volume was evident from 35 weeks onwards using micro-CT (Fig 5). There was some variability amongst different mice, with 2 out of 8 animals showing no detectable *Kras*$^{Lox\text{-}G12D}$ allele in blood at 40 weeks p.i. These same two mice were found to show low tumour volumes as determined by micro-CT. Thus, although these data demonstrate that ctDNA can be detected in the circulation of *Kras*$^{+/Lox\text{-}G12D}$ mice at extended time points, changes in tumour volume are not reliably detected earlier by ctDNA than by micro-CT.

DISCUSSION

The overall survival of cancer patients is greatly improved if the disease is diagnosed at an earlier stage, such as with CT screening in lung cancer (National Lung Screening Trial Research et al., 2013). However, the implementation of widespread CT screening in healthcare pathways is costly and impractical in many cases. Therefore, cheaper and more tractable alternatives are needed.

With regard to non-small cell lung cancer (NSCLC), extensive phylogenetic genomic sequencing in the TRACERx trial has identified evolutionary drivers of the disease (Jamal-Hanjani et al., 2017). Postoperative serial ctDNA profiling was shown to predict metastatic relapse in patients with NSCLC, and tumour volume, measured by CT volumetric analysis, correlated with the mean clonal plasma ctDNA variant allele frequency, becoming optimally detectable at volumes greater than 10 cm^3 (Abbosh et al., 2017).

The application of ctDNA monitoring to the earlier detection of lung cancer has been more difficult due to the challenges associated with the identification of patients with pre-malignant lesions. For this reason, we modelled this in a well-characterised genetic mouse model of lung adenocarcinoma development driven by

oncogenic KRAS (Jackson et al., 2001; Sutherland et al., 2014). One of the major strengths of the autochthonous *Kras*$^{LSL\text{-}G12D}$ mouse model is the ability to control tumour initiation by exposure to adenoviruses expressing Cre recombinase, coupled with the ability to follow subsequent tumour progression (Shaw et al., 2005). Mice develop numerous pulmonary lesions, which are predominantly of the papillary subtype, and these lesions include AAH, resembling putative precursors of human lung adenocarcinoma (Nikitin et al., 2004). These progress to small adenomas that enlarge over time, with some developing into overt adenocarcinomas over prolonged periods (Jackson et al., 2001). By monitoring the time from Cre induction, this model enables the investigation of mice with defined, early stage lung lesions. We show for the first time that cfDNA/ctDNA is released into the blood in this model and that cfDNA/ctDNA level correlates with tumour volume as determined by micro-CT. Furthermore, rising cfDNA levels and ctDNA were both detected without evidence of progression to overtly invasive adenocarcinomas suggesting that pre-malignant lung lesions are able to release fragments of DNA into the circulation.

Using micro-CT, we were able to serially track tumour growth, and we show that this approach provides an accurate volumetric measurement of lesions larger than 0.5 mm^3 by comparison to histological quantitation of tumour volumes (Fig 1E). Contrast agents have been used previously in the *Kras*$^{+/LSL\text{-}G12D}$ mouse model to better differentiate tumours from the surrounding vasculature (Lalwani et al., 2013), and detection of smaller tumours has been enhanced using bioluminescence and fluorescence biomarkers (Rodriguez et al., 2014). The use of these additional approaches and/or the implementation of more advanced, high-resolution techniques such as PET or MRI would potentially facilitate the detection of individual tumours with volumes below 0.5 mm^3 in the mouse.

Detection of the *Kras*$^{Lox\text{-}G12D}$ allele in total cfDNA by PCR proved extremely challenging due to the palindromic nature of the LoxP sequence and sequence homology between the *Kras*^{WT} and *Kras*$^{Lox\text{-}G12D}$ alleles. In addition, data from Supplementary Figure 5 show that Cre-mediated recombination introduces mutations at the recombination site, which potentially affect the performance of the assay. After multiple attempts, using several different approaches, a specific ddPCR assay was developed that was applied to plasma cfDNA samples. To our knowledge, this represents the first report of assays for detection of the Cre-recombined *Kras*$^{Lox\text{-}G12D}$ allele in *Kras*$^{LSL\text{-}G12D}$ mice.

The comparison between mice infected with Ad5-CMV-Cre and those infected with Ad5-SPC-Cre shows that, at similar tumour burden, there are comparable total cfDNA levels in mice regardless of the adenovirus used (Fig 3). This observation suggests that the infection of lung resident myeloid lineage cells and/or the development of bronchial lesions by Ad5-CMV-Cre does not significantly impact on the release of cfDNA, but this requires further exploration.

Total levels of circulating cfDNA are known to increase in patients with advanced disease (Leon et al., 1977; Madhavan et al., 2014; Newman et al., 2014). Consistently, we observed an increase in the total cfDNA concentration in mice with progressive disease (Fig 3). However, the relationship between tumour size and cfDNA concentration in mice was driven primarily by mice with significantly larger-than-average tumour burden. Notably, elevated cfDNA/ctDNA was detected in mice with at least 14 tumours larger than 0.5 mm^3, combining to a total volume above 7 mm^3. For a single mass this is equivalent to a tumour with diameter ~2.5 mm. Although, in humans, CT is able to detect lung nodules as small as 1 mm in diameter, the malignant potential of such lesions is unknown and follow up scans are

recommended to monitor progression of suspicious lesions (Rubin, 2015). Potentially, cfDNA/ctDNA profiling could be implemented at this stage to monitor the emergence of larger pre-malignant or indeed malignant lesions and thus avoid multiple CT scanning. However, it should be borne in mind that the volumes of mouse plasma analysed in the present study (~50 μl) relative to total mouse blood volume (2 ml) are not achievable in humans (typically 2 ml from a total blood volume of 5 litres) and therefore would require the development of more sensitive assays to profile the appropriate variant alleles.

In summary, we demonstrate the ability to detect the release of total cfDNA and tumour-derived ctDNA from mice bearing pre-malignant lung lesions with a total tumour volume in excess of 7.0 mm^3. This discovery is encouraging for the use of cfDNA/ctDNA profiling in the detection of pre-malignant lesions of the lung adenocarcinoma pathway in humans.

MATERIALS AND METHODS

Animal husbandry

Animal experiments were performed under the UK Home Office (HO) licence authority. Infected mice analysed at experimental endpoints underwent regulated procedures with a maximum severity classification of “moderate” according to the HO guidelines. *Kras*$^{+/LSL\text{-}G12D}$ mice were genotyped according to the Jacks laboratory recommended protocol (https://jacks-lab.mit.edu/protocols/genotyping/kras_cond). Intranasal inhalation of adenoviral vectors was performed using Ad5-CMV-Cre, Ad5-mSPC-Cre or Ad5-CMV-βgal in mice with an age range between 8 and 20 weeks. Mice were anaesthetised with 3% vapourised isofluorane in oxygen, in an induction

chamber. The viral supernatant (50 μL) was loaded into the nasal aperture of each individual animal. Viral concentrations were $5x10^7$ PFUs for Ad5-CMV-Cre and Ad5-CMV-βgal and $1x10^8$ for Ad5-mSPC-Cre viruses. All packaged adenoviruses were purchased from the Viral Vector Core Facility (University of Iowa, USA).

Blood collection and plasma isolation

For longitudinal time points, blood was withdrawn from saphenous veins, whereas cardiac blood samples were taken at the endpoint. For mouse saphenous blood sampling, 40 µL was pipetted into 200 µL EDTA (4.5 mM, pH 8.0) in phosphate-buffered saline (PBS). For cardiac blood sampling, blood (typically, 200 µL) was collected from terminally anaesthetised mice into K_3-EDTA vacutainers (BD Biosciences). Blood was centrifuged at 1000 x g for 10 minutes, and the plasma supernatant was centrifuged at 1000 x g for a further 10 minutes. Blood samples were processed promptly to reduce haemolysis; samples with evidence of haemolysis by visual inspection were excluded from further analysis. Plasma volumes were adjusted to 200 µL with PBS before extraction of cfDNA.

Extraction of circulating-free DNA

cfDNA was extracted from 200 µL plasma sample using the QIAamp DNA Blood Mini kit (Qiagen) and eluted in 50 µL TE buffer (10.0 mM Tris, 0.5 mM EDTA, pH 8.0). Purified DNA was stored at -20°C prior to use. Once extracted, the cfDNA concentration was measured using a Qubit 2.0 fluorometer with Qubit dsDNA HS Assay reagents (Thermo Fisher Scientific), with a detection range of ≥ 10 pg/µL.

Micro-CT imaging

Lung and tumour volumes were quantified using a Quantum FX micro-CT Imaging System (PerkinElmer). Animals were anaesthetised using a continuous flow of vapourised isofluorane. Animals were imaged for 34 seconds using a field-of-view (FOV) of 40 mm with respiratory gating. Each of these scans subjected the animal to ~20 mGy radiation, a relatively low dose well within the range recommended to avoid irradiation artefacts over the course of the study. To determine volumes for identified tumours and whole lung volumes, Caliper micro-CT Analysis software was used.

Histological analysis of tumours

Lungs were resected and fixed in 4% paraformaldehyde for 24 hours at room temperature. Tissue was then transferred to 70% ethanol and stored at 4°C. Tissue was embedded in paraffin, and 5 μm sections were cut using a microtome. For serial sectioning, sections at 100 μm intervals throughout the entire block were obtained. After sectioning, slides were stained with haematoxylin and eosin as previously reported (Kamata et al., 2015). Stained slides were photographed using a Leica DM500 microscope and an ICC50 Camera (Leica), and overlapping images were merged using Adobe Photoshop (version 13.0.1.1). Slides were also scanned on a NanoZoomer-XR Digital slide scanner C12000 (Hamamatsu Photonics) and analysed using NDP.view2 software (Hamamatsu Photonics) to calculate tumour volumes.

Cell culture

Mouse embryonic fibroblasts (MEFs) derived from *Kras*$^{+/LSL\text{-}G12D}$ mouse embryos were cultured as previously described (Andreadi et al., 2012). MEFs were

infected with Ad5-CMV-Cre at a multiplicity of infection (MOI) of 500, and *Kras*$^{+/Lox\text{-}G12D}$ cells were grown by continuous culture. Genomic DNA from MEFs was extracted as previously described (Andreadi et al., 2012).

PCR analysis

Primers used for qPCR analysis of *Gapdh* in cfDNA were as follows: forward 5'-CCTCACAATCTGTCTCACCTTATT-3' and reverse 5'-GACCTCTGTAAGTCCGCTTTG-3'; a TaqMan probe with the sequence FAM-AGCCTTATTGTCCTCGGGCAT-BH1 was also used. For the *Kras*$^{Lox\text{-}G12D}$ assay, the following primers were used: forward 5'- CCAGTCAACAAAGAATACCGCAAGG -3' and reverse 5'- TCTGCATAGTACGCTATACCCTGTG -3'; a TaqMan probe with the sequence HEX- TCGACATAACTTCGTATA-BH1 was also used. Underlined nucleotides represent locked nucleic acid (LNA) bases.

Quantitative real-time PCR

Quantitative real-time PCR was performed as previously described (Rakhit et al., 2017) on the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific) using TaqMan Fast Universal PCR Master Mix (Thermo Fisher Scientific). Thermal cycling conditions were as follows: 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C (for the *Gapdh* assay) or 63°C (for the *Kras*$^{Lox\text{-}G12D}$ assay) for 20 seconds. Reactions were conducted in triplicate, including a no-template control and positive control. *Gapdh* levels were adjusted for the plasma volume used for each cfDNA extraction.

Droplet digital (dd) PCR

All reactions for ddPCR analysis were formulated as described previously (Rakhit et al., 2017). A QX200 Droplet Digital PCR System (Bio-Rad) was used, using the manufacturer's protocol and reagents. Thermal cycling conditions were as follows: 95°C for 10 minutes; 40 cycles of 95°C for 30 seconds and 63°C for 30 seconds; and 98°C for 10 minutes. The $Kras^{Lox\text{-}G12D}$ copies were adjusted for the plasma volume used for each cfDNA extraction. A no-template control and a positive control were included in every assay. Analysis was performed according to the manufacturer's instructions on QuantaSoft software (Bio-Rad).

Statistical analysis

Statistical analysis was performed using GraphPad Prism 6 (www.graphpad.com). The data are presented as the mean value, and the error bar indicates ± SD or ± SEM (as indicated). Significance is indicated as **** for $p < 0.0001$, *** for $p < 0.001$, ** for $p < 0.01$, and * for $p < 0.05$.

ACKNOWLEDGEMENTS

We are grateful to the staff at the Preclinical Research Facility at Leicester, in particular, to Justyna Janus for support with animal husbandry and CT analysis. We thank Julian Downward and David Hancock (CRUK Signal Transduction Laboratory) for sharing unpublished results. We also wish to posthumously thank Francois Lassailly (CRUK *in vivo* imaging facility) for assistance in setting up the CT analysis methodology.

COMPETING INTERESTS

The authors declare no competing or financial interests

AUTHOR CONTRIBUTIONS

Conceptualisation: J.A.S., C.P., L.M.M.; Methodology and investigation: R.M.T., C.P.R., M.K.; Data analysis: R.M.T., C.P.R., J. L. Q., J.A.S., C.P., L.M.M. Writing: all authors.

FUNDING

This work was funded by the Medical Research Council.

REFERENCES

Abbosh, C. Birkbak, N. J. Wilson, G. A. Jamal-Hanjani, M. Constantin, T. Salari, R. Le Quesne, J. Moore, D. A. Veeriah, S. Rosenthal, R. et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446-451.

Andreadi, C., Cheung, L. K., Giblett, S., Patel, B., Jin, H., Mercer, K., Kamata, T., Lee, P., Williams, A., McMahon, M. et al. (2012). The intermediate-activity (L597V)BRAF mutant acts as an epistatic modifier of oncogenic RAS by enhancing signaling through the RAF/MEK/ERK pathway. *Genes Dev* **26**, 1945-58.

Burrell, R. A., McGranahan, N., Bartek, J. and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338-45.

Cabral, R. E., Caldeira-de-Araujo, A., Cabral-Neto, J. B. and Costa Carvalho Mda, G. (2010). Analysis of GSTM1 and GSTT1 polymorphisms in circulating plasma DNA of lung cancer patients. *Mol Cell Biochem* **338**, 263-9.

Chaudhuri, A. A., Chabon, J. J., Lovejoy, A. F., Newman, A. M., Stehr, H., Azad, T. D., Khodadoust, M. S., Esfahani, M. S., Liu, C. L., Zhou, L. et al. (2017). Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. *Cancer discovery* **7**, 1394-1403.

Cheng, J., Cuk, K., Heil, J., Golatta, M., Schott, S., Sohn, C., Schneeweiss, A., Burwinkel, B. and Surowy, H. (2017). Cell-free circulating DNA integrity is an independent predictor of impending breast cancer recurrence. *Oncotarget* **8**, 54537-54547.

Diaz, L. A., Jr., Williams, R. T., Wu, J., Kinde, I., Hecht, J. R., Berlin, J., Allen, B., Bozic, I., Reiter, J. G., Nowak, M. A. et al. (2012). The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-40.

Esposito, D., Gillette, W. K. and Hartley, J. L. (2003). Blocking oligonucleotides improve sequencing through inverted repeats. *Biotechniques* **35**, 914-6, 918, 920.

Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R. J., Cheang, M., Osin, P., Nerurkar, A., Kozarewa, I. et al. (2015). Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med* **7**, 302ra133.

Goebel, G., Zitt, M., Zitt, M. and Muller, H. M. (2005). Circulating nucleic acids in plasma or serum (CNAPS) as prognostic and predictive markers in patients with solid neoplasias. *Dis Markers* **21**, 105-20.

Guerra, C., Mijimolle, N., Dhawahir, A., Dubus, P., Barradas, M., Serrano, M., Campuzano, V. and Barbacid, M. (2003). Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer cell* **4**, 111-20.

Huang, F. T., Yu, K., Balter, B. B., Selsing, E., Oruc, Z., Khamlichi, A. A., Hsieh, C. L. and Lieber, M. R. (2007). Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol Cell Biol* **27**, 5921-32.

Jackson, E. L., Olive, K. P., Tuveson, D. A., Bronson, R., Crowley, D., Brown, M. and Jacks, T. (2005). The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer research* **65**, 10280-8.

Jackson, E. L., Willis, N., Mercer, K., Bronson, R. T., Crowley, D., Montoya, R., Jacks, T. and Tuveson, D. A. (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* **15**, 3243-8.

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., Shafi, S., Johnson, D. H., Mitter, R., Rosenthal, R. et al. (2017). Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* **376**, 2109-2121.

Kamata, T., Giblett, S. and Pritchard, C. (2017). KRAS(G12D) expression in lung-resident myeloid cells promotes pulmonary LCH-like neoplasm sensitive to statin treatment. *Blood* **130**, 514-526.

Kamata, T., Jin, H., Giblett, S., Patel, B., Patel, F., Foster, C. and Pritchard, C. (2015). The cholesterol-binding protein NPC2 restrains recruitment of stromal macrophage-lineage cells to early-stage lung tumours. *EMBO Mol Med* **7**, 1119-37.

Kirsch, D. G., Grimm, J., Guimaraes, A. R., Wojtkiewicz, G. R., Perez, B. A., Santiago, P. M., Anthony, N. K., Forbes, T., Doppke, K., Weissleder, R. et al. (2010). Imaging primary lung cancers in mice to study radiation biology. *Int J Radiat Oncol Biol Phys* **76**, 973-7.

Lalwani, K., Giddabasappa, A., Li, D., Olson, P., Simmons, B., Shojaei, F., Van Arsdale, T., Christensen, J., Jackson-Fisher, A., Wong, A. et al. (2013). Contrast agents for quantitative microCT of lung tumors in mice. *Comp Med* **63**, 482-90.

Leon, S. A., Shapiro, B., Sklaroff, D. M. and Yaros, M. J. (1977). Free DNA in the serum of cancer patients and the effect of therapy. *Cancer research* **37**, 646-50.

Loonstra, A., Vooijs, M., Beverloo, H. B., Allak, B. A., van Drunen, E., Kanaar, R., Berns, A. and Jonkers, J. (2001). Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells. *Proc Natl Acad Sci U S A* **98**, 9209-14.

Madhavan, D., Wallwiener, M., Bents, K., Zucknick, M., Nees, J., Schott, S., Cuk, K., Riethdorf, S., Trumpp, A., Pantel, K. et al. (2014). Plasma DNA integrity as a biomarker for primary and metastatic breast cancer and potential marker for early diagnosis. *Breast Cancer Res Treat* **146**, 163-74.

Mandel, P. and Metais, P. (1948). Les acides nucleiques du plasma sanguin chez l'homme. *Comptes rendus des seances de la Societe de biologie et de ses filiales* **142**, 241-3.

National Lung Screening Trial Research, T., Church, T. R., Black, W. C., Aberle, D. R., Berg, C. D., Clingan, K. L., Duan, F., Fagerstrom, R. M., Gareen, I. F., Gierada, D. S. et al. (2013). Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med* **368**, 1980-91.

Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L., Neal, J. W., Wakelee, H. A., Merritt, R. E. et al. (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* **20**, 548-54.

Nikitin, A. Y., Alcaraz, A., Anver, M. R., Bronson, R. T., Cardiff, R. D., Dixon, D., Fraire, A. E., Gabrielson, E. W., Gunning, W. T., Haines, D. C. et al. (2004). Classification of proliferative pulmonary lesions of the mouse:

571 recommendations of the mouse models of human cancers consortium. *Cancer*
572 *research* **64**, 2307-16.

573 **Rakhit, C. P., Ottolini, B., Jones, C., Pringle, J. H., Shaw, J. A. and**
574 **Martins, L. M.** (2017). Peptide nucleic acid clamping to improve the sensitivity of Ion
575 Torrent-based detection of an oncogenic mutation in KRAS. *matters* **in press**.

576 **Rodriguez, E., Mannion, L., D'Santos, P., Griffiths, M., Arends, M. J.,**
577 **Brindle, K. M. and Lyons, S. K.** (2014). Versatile and enhanced tumour modelling
578 in mice via somatic cell transduction. *J Pathol* **232**, 449-57.

579 **Rubin, G. D.** (2015). Lung nodule and cancer detection in computed
580 tomography screening. *J Thorac Imaging* **30**, 130-8.

581 **Schwarzenbach, H., Hoon, D. S. and Pantel, K.** (2011). Cell-free nucleic
582 acids as biomarkers in cancer patients. *Nat Rev Cancer* **11**, 426-37.

583 **Shaw, A. T., Kirsch, D. G. and Jacks, T.** (2005). Future of early detection of
584 lung cancer: the role of mouse models. *Clin Cancer Res* **11**, 4999s-5003s.

585 **Sheridan, C. and Downward, J.** (2015). Overview of KRAS-Driven
586 Genetically Engineered Mouse Models of Non-Small Cell Lung Cancer. *Curr Protoc*
587 *Pharmacol* **70**, 14 35 1-16.

588 **Stroun, M., Anker, P., Lyautey, J., Lederrey, C. and Maurice, P. A.** (1987).
589 Isolation and characterization of DNA from the plasma of cancer patients. *Eur J*
590 *Cancer Clin Oncol* **23**, 707-12.

591 **Sutherland, K. D., Song, J. Y., Kwon, M. C., Proost, N., Zevenhoven, J.**
592 **and Berns, A.** (2014). Multiple cells-of-origin of mutant K-Ras-induced mouse lung
593 adenocarcinoma. *Proc Natl Acad Sci U S A* **111**, 4952-7.

Trigg, R. M. (2017). Molecular analysis of circulating cell-free DNA in lung cancer. In *Cancer Studies and Molecular Medicine*, vol. PhD. Leicester Research Archive: University of Leicester.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15.

FIGURE LEGENDS

Figure 1. Comparison of micro-CT imaging and histological analysis of tumours for measuring tumour burden.

(**A**) Schematic of the *Kras*$^{LSL\text{-}G12D}$ allele. Expression of Cre recombinase allows for removal of the LoxP-STOP-LoxP cassette and formation of the *Kras*$^{Lox\text{-}G12D}$ allele expressing Kras^{G12D}. The anatomy of the mouse lung is shown in (i). Tumour number and size were estimated at the endpoint by micro-CT (ii, tumours in red) or histology (iii) following 40 weeks of induction of mice with Ad5-mSPC-Cre. The histology image shows tumours with different volumes.

(**B**) Comparison of the total number of tumours detected by micro-CT and histology in 5 individual mice at the endpoint 40 weeks p.i. with Ad5-mSPC-Cre. There were significantly more tumours ($p < 0.01$) detected by histological analysis than by micro-CT (mean ± SD; asterisks, paired *t*-test).

(**C**) Comparison of the total number of tumours at endpoint (40 weeks p.i. with Ad5-mSPC-Cre), detected by micro-CT and histology, grouped (binned) according to size, in 5 individual animals (mean ± SD; asterisk, paired *t*-test).

(**D**) Size distribution of tumours detected by micro-CT (top row) or histology (bottom row) in 5 individual mice at endpoint (40 weeks p.i. with Ad5-mSPC-Cre). The total number of tumours identified by the two methods in each animal (n) is indicated.

(**E**) Comparison of tumour volumes as determined by micro-CT and serial sectioning of the entire tumour, H&E staining and tumour area quantitation. The two methods showed a linear correlation.

Figure 2. Histological assessment of pathological alterations in mice expressing the *Kras$^{Lox\text{-}G12D}$* allele.

(**A**) Representative H&E-stained lung sections of *Kras$^{+/LSL\text{-}G12D}$* mice infected with Ad5-CMV-Cre expression and analysed 20 weeks post-infection. The different pathologies detected in the far left image are magnified in images i-iii showing representative examples of alveolar hyperplasia (i), papillary adenoma (ii) and papillary endobronchial lesions (iii).

(**B**) Quantitative assessment of each pathology detected in H&E-stained sections of lung tissue collected from *Kras$^{+/LSL\text{-}G12D}$* mice infected with either Ad5-CMV-Cre (n 4) or Ad5-mSPC-Cre (n = 3) at endpoint (20 and 40 weeks, respectively).

Figure 3. Monitoring of cfDNA levels using *Gapdh* analysis by qPCR.

(A) *Gapdh* levels as measured by qPCR in cfDNA in comparison to tumour burden in *Kras$^{+/LSL\text{-}G12D}$* mice over a time course following infection with Ad5-CMV-Cre (red) or Ad5-CMV-βgal (blue). Total tumour burden was quantitated from micro-CT imaging. *Gapdh* levels were measured by qPCR of circulating DNA derived from the plasma of mice at each time point. Mean values are indicated by diamonds/lines, while values for individual mice are indicated by circles (*n* = 3-12 at each time point). The asterisks indicate $p < 0.05$ for two-tailed unpaired *t*-test comparisons between mean values at a given time point and values at time = 0. The correlation coefficients between *Gapdh* levels and tumour burden are indicated (linear regression analysis; *p* value, goodness of fit).

(B) *Gapdh* levels in cfDNA as measured by qPCR in comparison to tumour burden in *Kras$^{+/LSL\text{-}G12D}$* mice over a time course following infection with Ad5-mSPC-Cre. Mean values are indicated by diamonds/lines, while values for individual mice are

650 indicated by circles (n = 4-9 at each time point). The asterisks indicate $p < 0.05$ for
651 two-tailed unpaired t-test comparisons between mean values at a given time point
652 and values at time = 0. The correlation coefficients between *Gapdh* levels and
653 tumour burden are indicated (linear regression analysis; p value, goodness of fit).
654
655 **Figure 4. Validation of ddPCR assay for detection of *Kras*$^{Lox\text{-}G12D}$ allele using**
656 **LNA-2 primers.**
657 One-dimensional droplet plot of LNA-2 assay (comprising probe LNA-2 and primers
658 flanking the recombined LoxP sequence in *Kras*$^{Lox\text{-}G12D}$) with (**A**) *Kras*$^{+/Lox\text{-}G12D}$ MEF
659 genomic DNA; (**B**) *Kras*$^{+/LSL\text{-}G12D}$ MEF genomic DNA; and (**C**) *Kras*$^{+/+}$ MEF genomic
660 DNA. Primers were annealed at 62°C, following a gradient PCR experiment to
661 determine the optimum annealing temperature (T_a, not shown). A manual threshold
662 of 1,000 fluorescence units was selected.
663 (**D**) *Kras*$^{+/Lox\text{-}G12D}$ MEF genomic DNA was serially diluted into a background of
664 *Kras*$^{+/LSL\text{-}G12D}$ MEF genomic DNA. A one-dimensional droplet plot for the 133-bp
665 LNA-2 assay at each serial dilution is shown. A manual threshold of 1,100
666 fluorescence units was selected.
667 (**E**) Actual vs. theoretical copy number of *Kras*$^{Lox\text{-}G12D}$ at each serial dilution. Error
668 bars represent the mean ± SD.
669
670 **Figure 5. Detection of the *Kras*$^{Lox\text{-}G12D}$ allele in plasma using ddPCR.**
671 Analysis of *Kras*$^{Lox\text{-}G12D}$ levels in cfDNA by ddPCR and the total tumour burden in
672 *Kras*$^{+/LSL\text{-}G12D}$ mice over a time course following infection with Ad5-mSPC-Cre.
673 Individual values are indicated by circles. Mean values are indicated by
674 diamonds/lines, while values for individual mice are indicated by circles (n = 8-9 at

each time point). The asterisks indicate $p < 0.05$ for two-tailed unpaired *t*-test comparisons between mean values at a given time point and values at time = 0. The correlation coefficients between $Kras^{Lox\text{-}G12D}$ levels and tumour burden are indicated (linear regression analysis; *p* value, goodness of fit). Datasets for total tumour volume are also used in Figure 3.

SUPPLEMENTARY INFORMATION

Supplementary Table 1.

Theoretical copy number and frequency of *Kras*$^{Lox\text{-}G12D}$ in each serial dilution for the sensitivity analysis of the 133 bp LNA-2 assay

Serial dilution	*Kras*$^{+/Lox\text{-}G12D}$ gDNA (pg)	*Kras*$^{+/LSL\text{-}G12D}$ gDNA (pg)	*Kras*$^{Lox\text{-}G12D}$ copy number	*Kras*$^{Lox\text{-}G12D}$ frequency (%)
1	2500	5000	379	33.0
2	1250	5000	189	20.0
3	625	5000	95	11.0
4	312.5	5000	47	5.9
5	156.3	5000	24	3.0
6	78.1	5000	12	1.6
7	39.1	5000	6	0.8
8	19.5	5000	3	0.4
9	9.8	5000	1.5	0.2
10	4.9	5000	0.75	0.1
11	2.4	5000	0.38	0.05

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. LoxP sequence and putative stem-loop structures.

(**A**) LoxP is a 34-bp sequence comprising two 13-bp palindromic sequences flanking a central 8-bp spacer sequence.

(**B**) LoxP forms a predicted 13-bp stem-loop structure with a melting temperature (T_m) of 55.6°C.

(**C**) In the *Kras*$^{LSL\text{-}G12D}$ allele, Cre-mediated recombination leads to the formation of a predicted 19-bp stem-loop structure with a T_m of 65.5°C. The 6-bp *Sal*1 restriction site is shown in grey, thus extending the stem-loop by 6 bp in the *Kras*$^{Lox\text{-}G12D}$ allele.

Supplementary Figure 2. Schematic diagram of the TaqMan and locked nucleic acid-modified probes targeting the recombined LoxP sequence in *Kras*$^{Lox\text{-}G12D}$.

Probes MGB1-3 are TaqMan 5' FAM, 3' MGB-labelled hydrolysis probes targeting various regions of LoxP/*Sal*1 on the antisense strand. Probes LNA1-2 are 5' HEX, 3' BHQ1-labelled probes with some bases 'locked' (black circles) using locked nucleic acid (LNA) technology. LNA-1 targets LoxP/*Sal*1 on the antisense strand only, whereas LNA-2 targets LoxP/*Sal*1 on both the sense and antisense strands.

Supplementary Figure 3. Performance of the 113-bp LNA-1 assay by qPCR.

(**A**) qPCR amplification plot of probe LNA-1 and primers flanking the recombined LoxP sequence, using a template of 100 ng of *Kras*$^{+/Lox\text{-}G12D}$ MEF gDNA.

(**B**) Performance of standard curves of serially diluted *Kras*$^{+/Lox\text{-}G12D}$ MEF gDNA in the absence (solid line) and presence (dashed line) of *Kras*$^{+/LSL\text{-}G12D}$ MEF gDNA. In a

background of *Kras*$^{+/LSL\text{-}G12D}$ gDNA, the PCR signal was lost after the fourth dilution, corresponding to a *Kras*$^{Lox\text{-}G12D}$ frequency of 6.25%.

Supplementary Figure 4. Performance of the 113-bp LNA-1 assay by ddPCR.
One-dimensional droplet plot of the 113-bp LNA-1 assay (comprising probe LNA-1 and primers flanking the recombined LoxP sequence in *Kras*$^{Lox\text{-}G12D}$) with (**A**) *Kras*$^{+/Lox\text{-}G12D}$ MEF gDNA; (**B**) *Kras*$^{+/LSL\text{-}G12D}$ MEF gDNA; and (**C**) *Kras*$^{+/+}$ MEF gDNA. A manual threshold of 5,000 fluorescence units was selected. Primers were annealed at 62°C, following a gradient PCR experiment to determine the optimum T_a (not shown). Red arrowed lines indicate areas of 'rain' caused by negative droplets above the main negative droplet clusters.

Supplementary Figure 5. Assessment of sequence variation following Cre-mediated recombination
(**A**) H&E-stained section to indicate location of core biopsy from *Kras*$^{+/Lox\text{-}G12D}$ mouse lung. A 1.5-mm-diameter core biopsy was punched from a region of the *Kras*$^{+/Lox\text{-}G12D}$ mouse lung tissue, which was confirmed by H&E analysis to contain multiple adenomas.
(**B**) Alignment of reads for sequences within and around the recombined LoxP sequence in *Kras*$^{Lox\text{-}G12D}$; 140/180-bp primers flanking the recombined LoxP sequence were used to amplify a 180-bp sequence in *Kras*$^{Lox\text{-}G12D}$. Mutations are shown in yellow. Primer sequences are not shown.
(**C**) A schematic diagram showing the position of mutations within and around the recombined LoxP sequence (black) in *Kras*$^{Lox\text{-}G12D}$. The grey boxes show the positions of the primers and LNA-2 probe in the 133-bp LNA-2 assay.

SUPPLEMENTARY METHODS

Quantitative PCR with LNA bases

An alternative probe-based strategy was utilized in which a hydrolysis probe was designed to include so-called 'locked nucleic acid' (LNA) bases, a chemical modification that increases both the T_m and binding specificity of the probe while also providing strand-invasion properties (Suppl. Fig 2). This LNA probe, LNA-1, permitted PCR amplification of the recombined allele with a template of *Kras*$^{+/Lox\text{-}G12D}$ gDNA, but efficiency was poor (~ 85%). Moreover, generation of a standard curve of *Kras*$^{+/Lox\text{-}G12D}$ gDNA diluted into a background of *Kras*$^{+/LSL\text{-}G12D}$ (unrecombined allele) showed a complete dropout of the qPCR signal after the fourth dilution, corresponding to a recombined allele frequency of 6.25% (Suppl. Fig 3). Therefore, the LNA-1 probe strategy was unable to detect *Kras*$^{Lox\text{-}G12D}$ at low allele frequencies.

PCR amplification, cloning and sequencing of DNA of lung tissue

Sequence analysis of LoxP sequences following Cre-mediated recombination of the *Kras*$^{LSL\text{-}G12D}$ allele has not been reported previously. Since the assay for the detection of the Cre-Lox recombination requires the annealing of a probe to the LoxP sequence, mutations in and around LoxP may compromise assay performance. To investigate possible sequence variations, DNA was extracted from an FFPE core biopsy of *Kras*$^{+/Lox\text{-}G12D}$ mouse lung tissue, which was confirmed to contain multiple adenomas by haematoxylin and eosin histological analysis (Suppl. Fig 5A). A 180-bp region surrounding the recombined LoxP sequence was PCR amplified and cloned into a vector. Clones were screened by PCR, and 40 clones containing the 180-bp insert were sequenced. Sequence alignment revealed point mutations within 12 of the 40 clones containing recombined LoxP (Suppl. Fig 5B). These mutations

clustered within and around the *Sal*1 restriction sites and occurred in intronic sequences (Suppl. Fig 5C). Eleven of the 12 mutations were transitions (4 A>G; 1 G>A; 4 T>C; 2 C>T), and all mutations were confirmed by sequencing the opposite strand. To confirm that these mutations were not induced by errors from the DNA polymerase during PCR amplification of the insert sequence, the 140-bp WT region of *Kras* (co-amplified by the same primers) was also cloned. Forty clones were sequenced, and no mutations were identified (data not shown), thus confirming that the mutations observed following Cre-mediated recombination of LoxP sequences were induced by the recombination process.

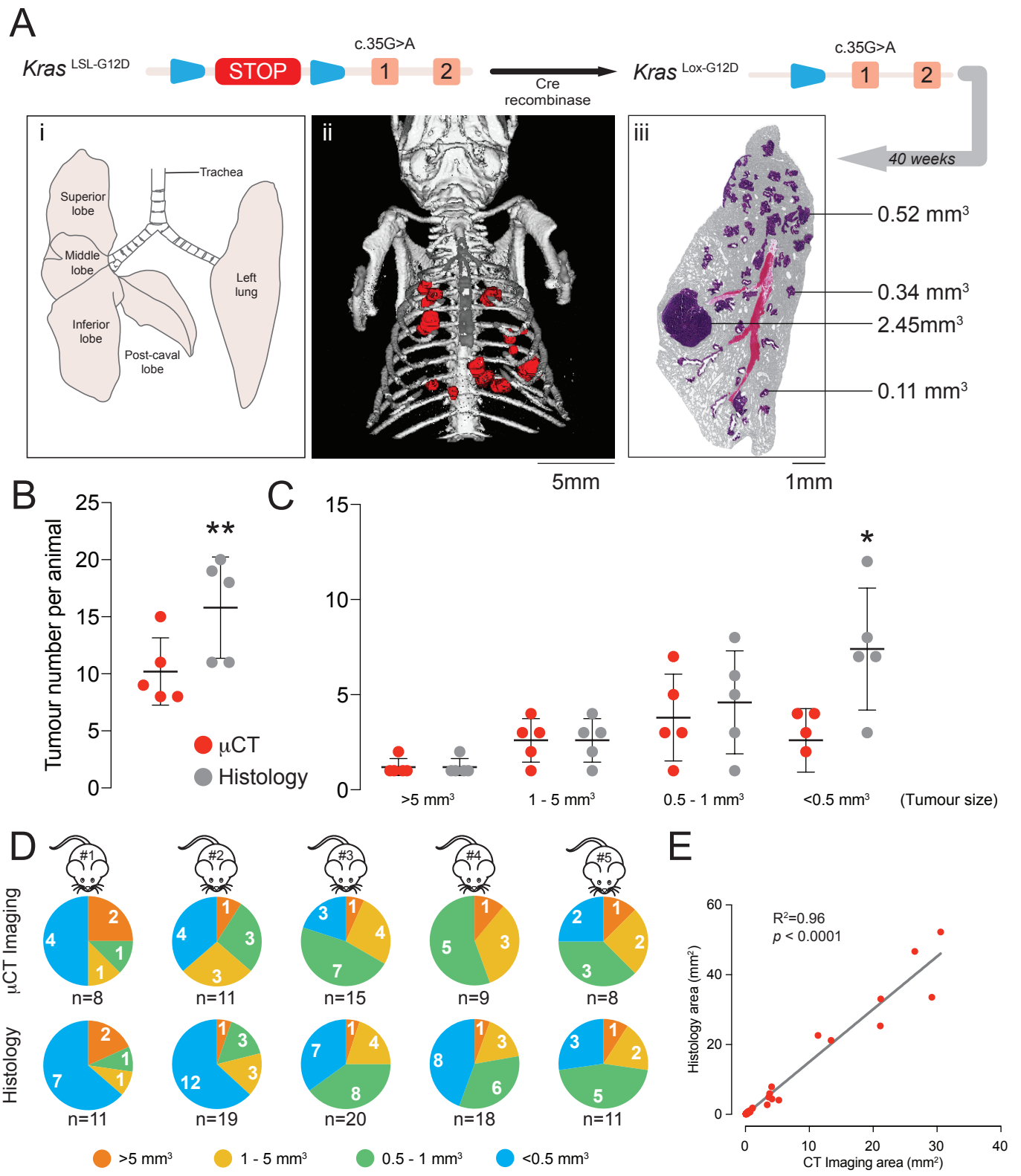

A
Kras LSL-G12D
STOP
c.35G>A
1
2
Cre recombinase
Kras Lox-G12D
c.35G>A
1
2
40 weeks
i
Trachea
Superior lobe
Middle lobe
Left lung
Inferior lobe
Post-caval lobe
ii
5mm
iii
0.52 mm³
0.34 mm³
2.45mm³
0.11 mm³
1mm
B
Tumour number per animal
25
20
15
10
5
0
**
μCT
Histology
C
15
10
5
0
*
>5 mm³
1 - 5 mm³
0.5 - 1 mm³
<0.5 mm³
(Tumour size)
D
μCT Imaging
Histology
#1
#2
#3
#4
#5
n=8
n=11
n=15
n=9
n=8
n=11
n=19
n=20
n=18
n=11
>5 mm³
1 - 5 mm³
0.5 - 1 mm³
<0.5 mm³
E
R²=0.96
p < 0.0001
Histology area (mm²)
CT Imaging area (mm²)
0
10
20
30
40
60

Figure 1

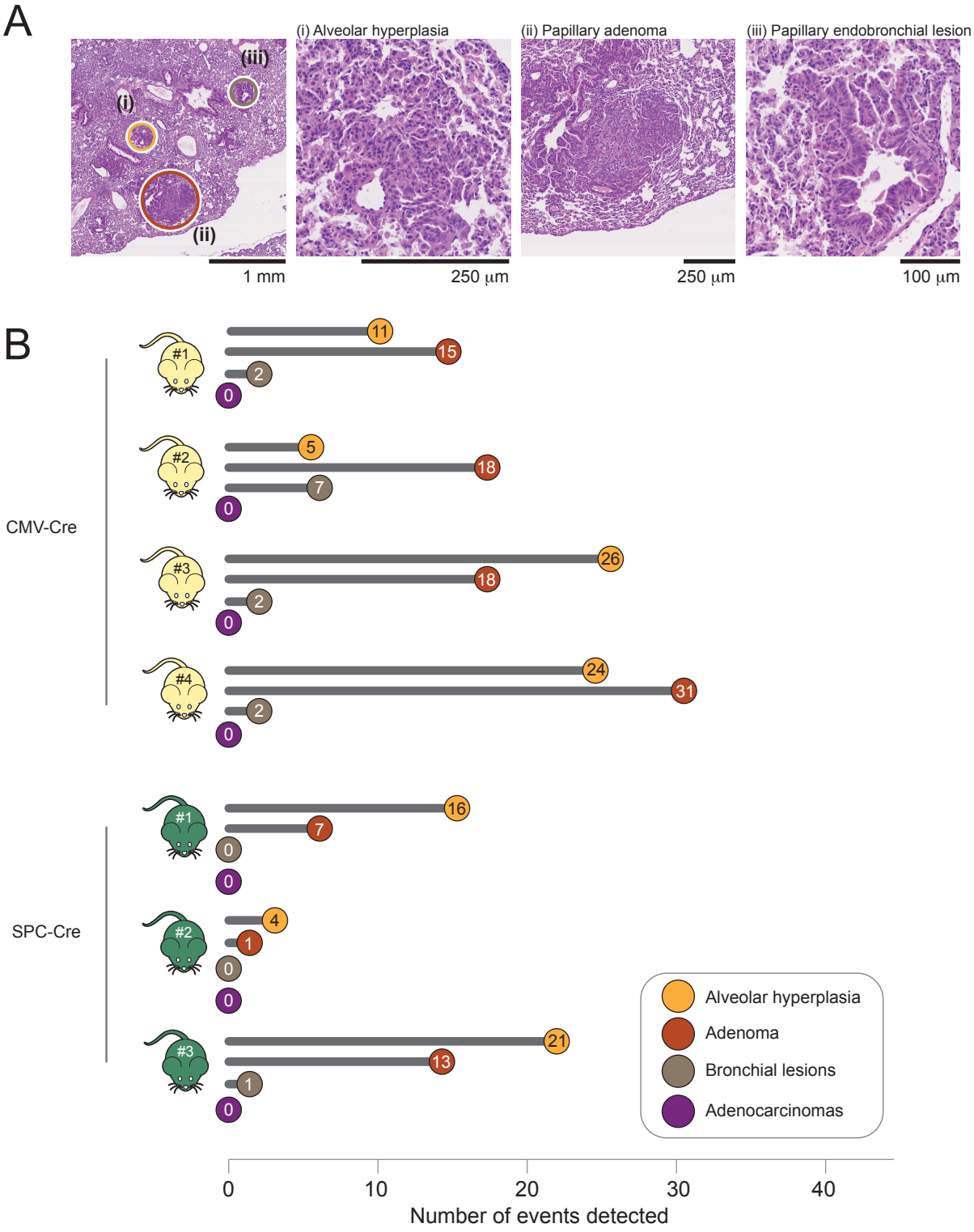

A
(i) Alveolar hyperplasia
(ii) Papillary adenoma
(iii) Papillary endobronchial lesion
(iii)
(i)
(ii)
1 mm
250 μm
250 μm
100 μm
B
CMV-Cre
SPC-Cre
#1
#2
#3
#4
11
15
2
0
5
18
7
0
26
18
2
0
24
31
2
0
16
7
0
0
4
1
0
0
21
13
1
0
Alveolar hyperplasia
Adenoma
Bronchial lesions
Adenocarcinomas
0
10
20
30
40
Number of events detected

Figure 2

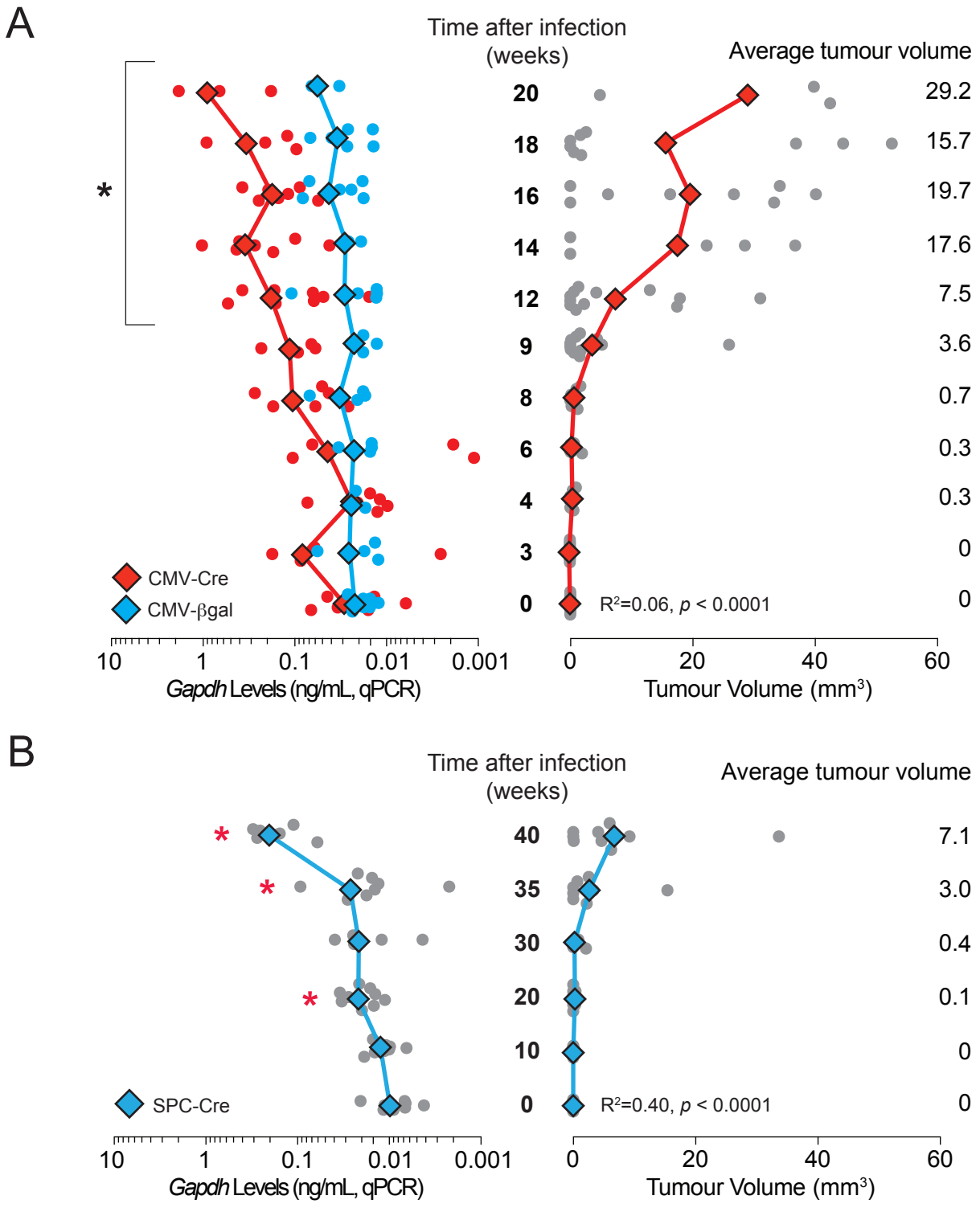

A
Time after infection (weeks)
Average tumour volume
20
18
16
14
12
9
8
6
4
3
0
29.2
15.7
19.7
17.6
7.5
3.6
0.7
0.3
0.3
0
0
*
CMV-Cre
CMV-βgal
R²=0.06, p < 0.0001
10
1
0.1
0.01
0.001
Gapdh Levels (ng/mL, qPCR)
0
20
40
60
Tumour Volume (mm³)
B
Time after infection (weeks)
Average tumour volume
40
35
30
20
10
0
7.1
3.0
0.4
0.1
0
0
*
*
*
SPC-Cre
R²=0.40, p < 0.0001
10
1
0.1
0.01
0.001
Gapdh Levels (ng/mL, qPCR)
0
20
40
60
Tumour Volume (mm³)

Figure 3

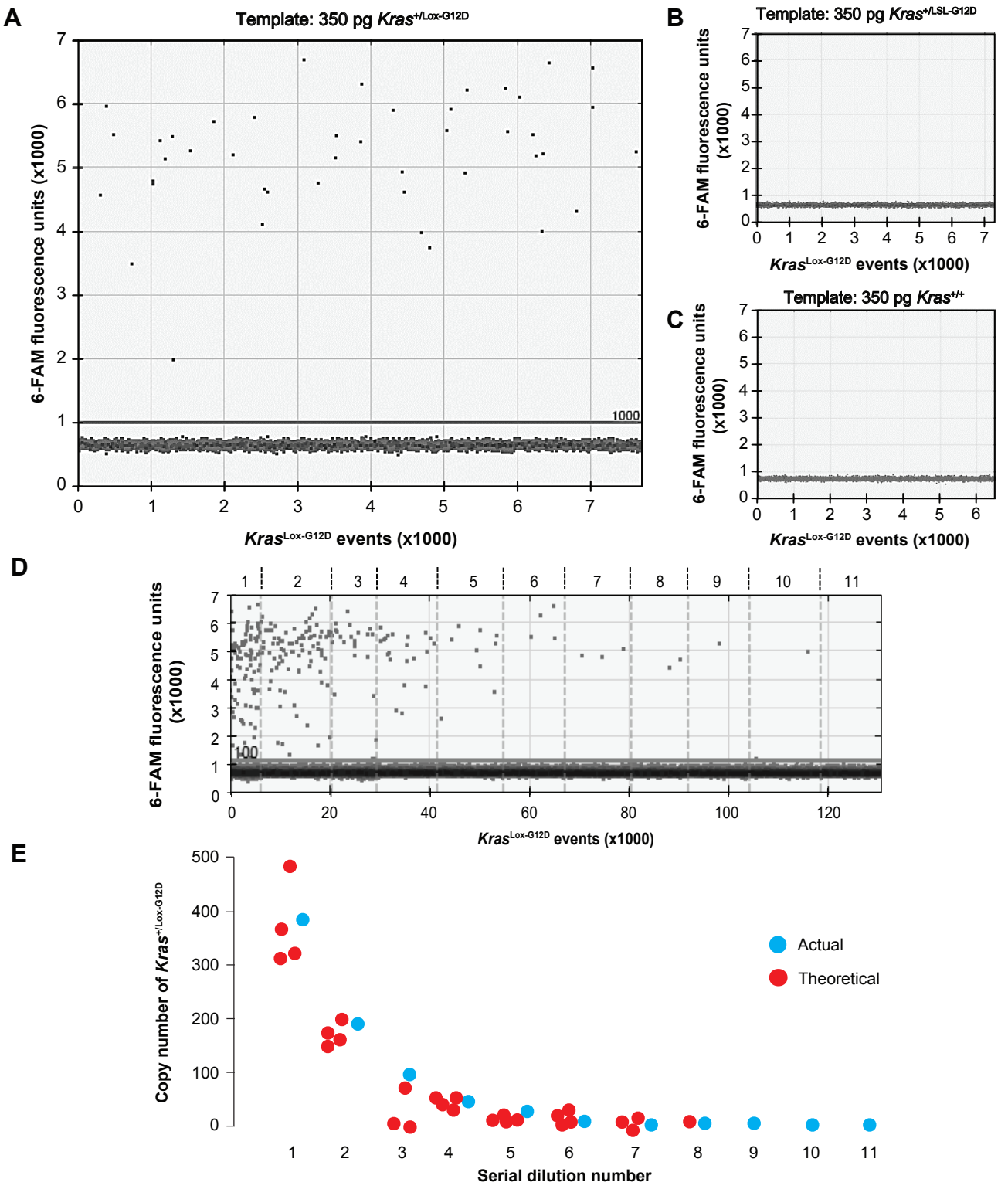

A
Template: 350 pg Kras+/Lox-G12D
6-FAM fluorescence units (x1000)
1000
KrasLox-G12D events (x1000)
B
Template: 350 pg Kras+/LSL-G12D
6-FAM fluorescence units (x1000)
KrasLox-G12D events (x1000)
C
Template: 350 pg Kras+/+
6-FAM fluorescence units (x1000)
KrasLox-G12D events (x1000)
D
6-FAM fluorescence units (x1000)
100
KrasLox-G12D events (x1000)
E
Copy number of Kras+/Lox-G12D
Actual
Theoretical
Serial dilution number

Figure 4

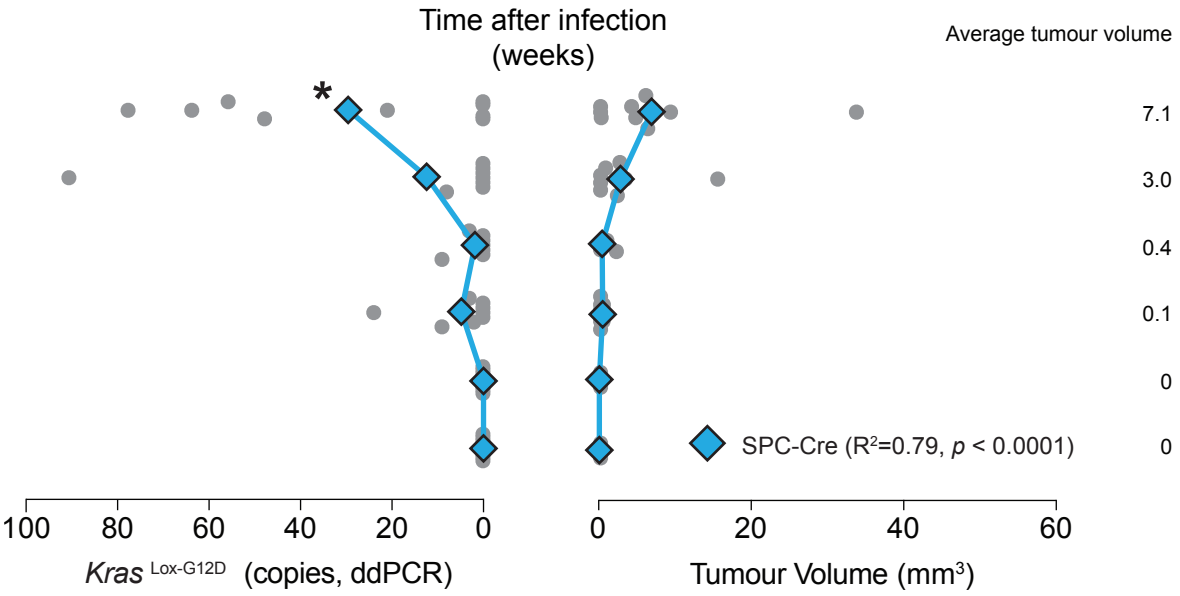
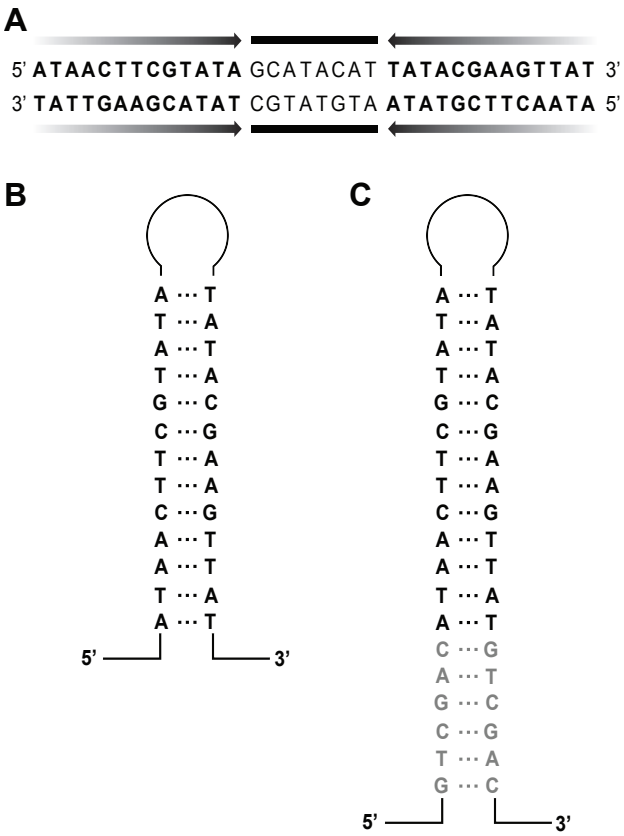
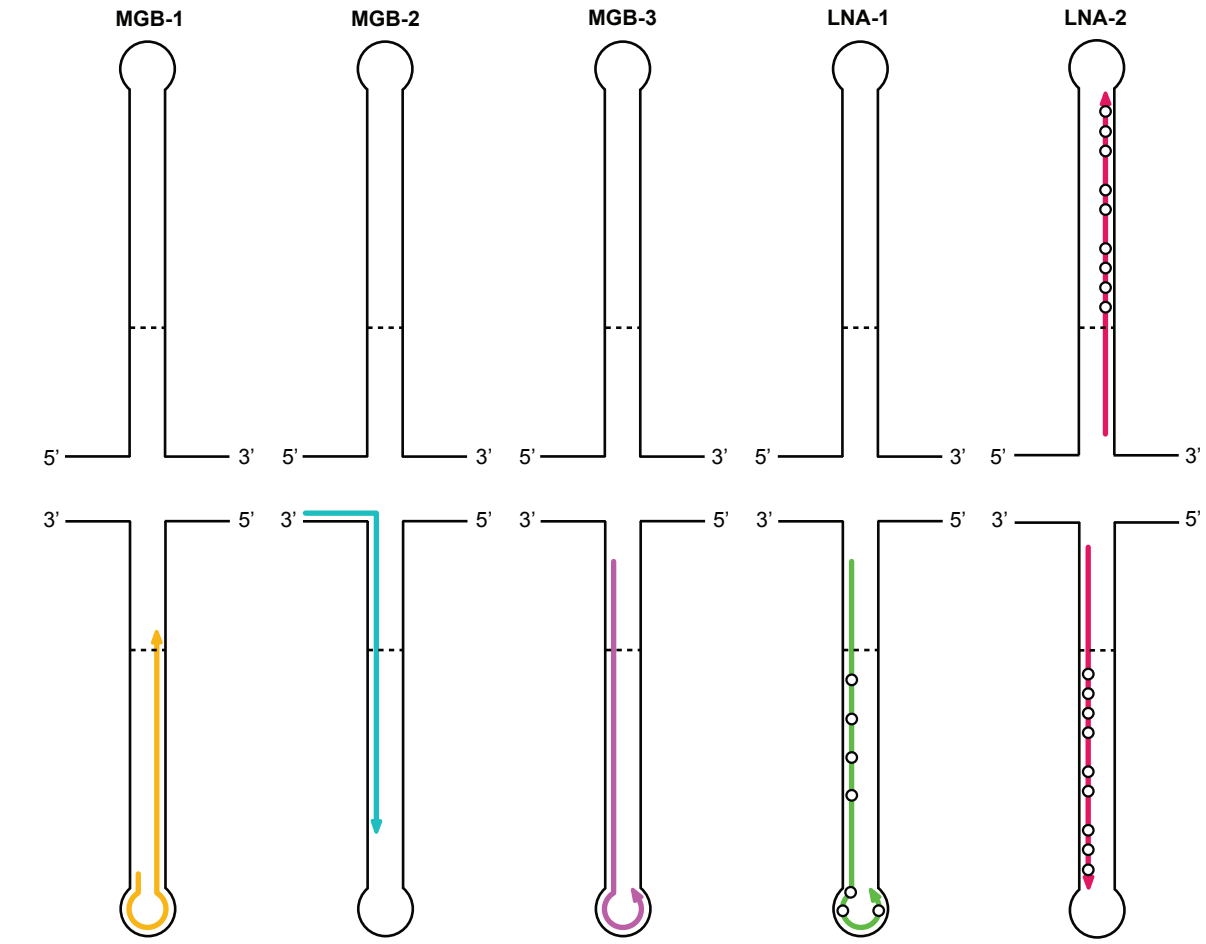

Time after infection
(weeks)
Average tumour volume
7.1
3.0
0.4
0.1
0
0
*
SPC-Cre (R^2=0.79, p < 0.0001)
100
80
60
40
20
0
Kras Lox-G12D (copies, ddPCR)
0
20
40
60
Tumour Volume (mm³)

Figure 5

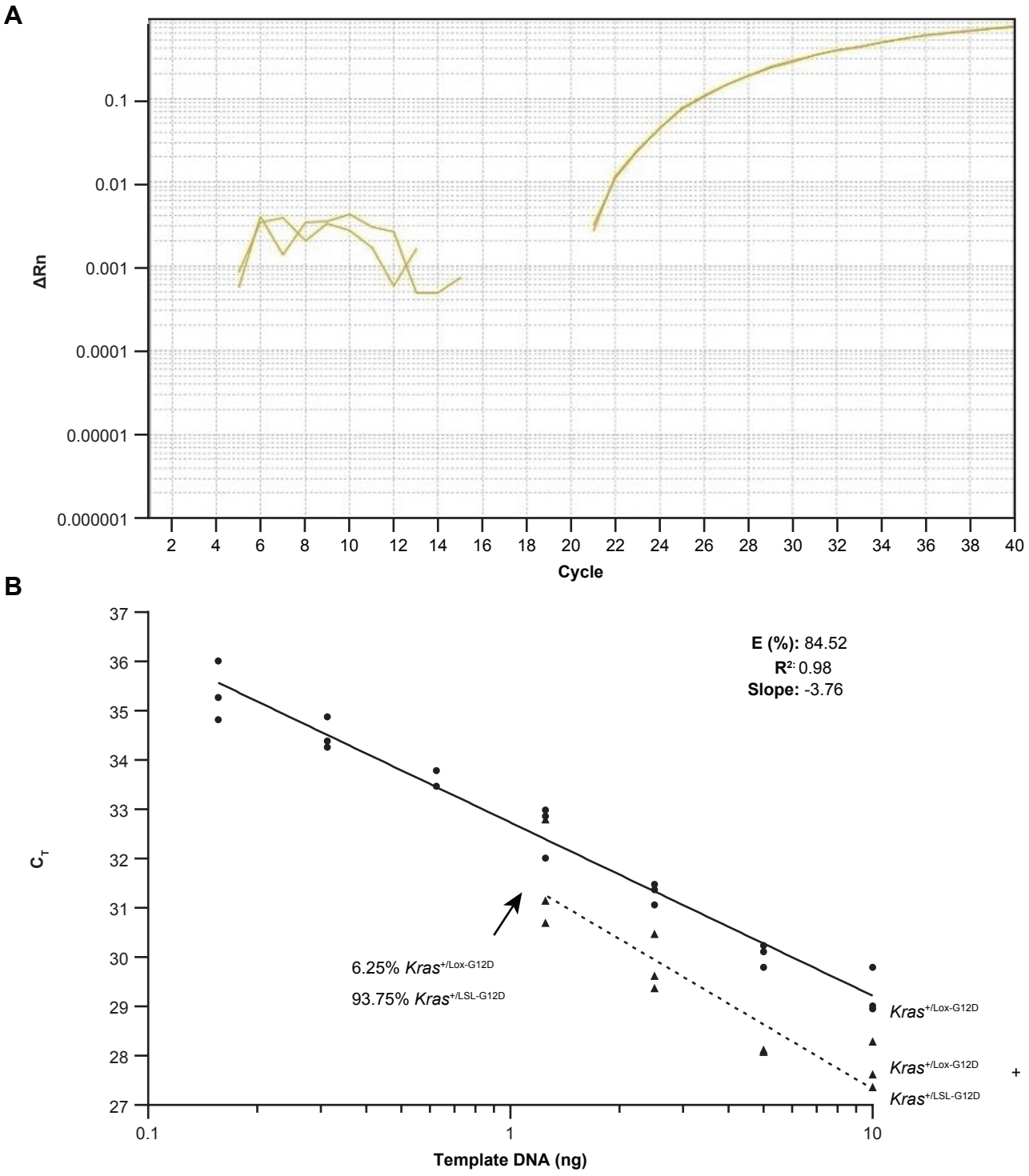


Supplementary Figure 1

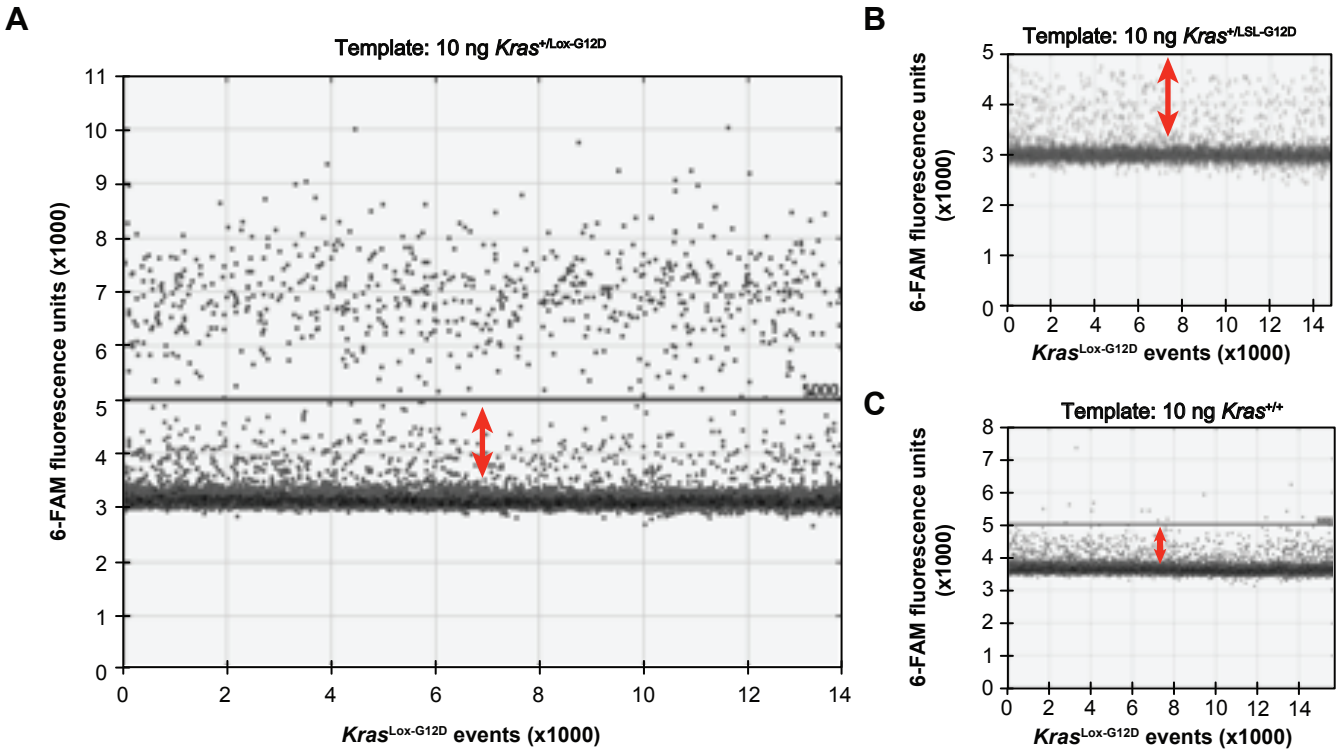


	MGB-1	MGB-2	MGB-3	LNA-1	LNA-2
Tm (°C)	70	74	72	70	74
Length	22	20	25	24	18
LNA	0	0	0	7	9
Strand	AS	AS	AS	AS	AS/S

Supplementary Figure 2

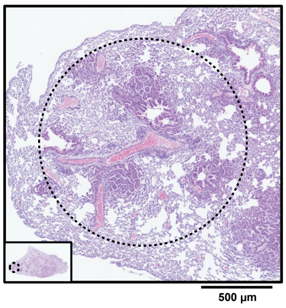


Supplementary Figure 3

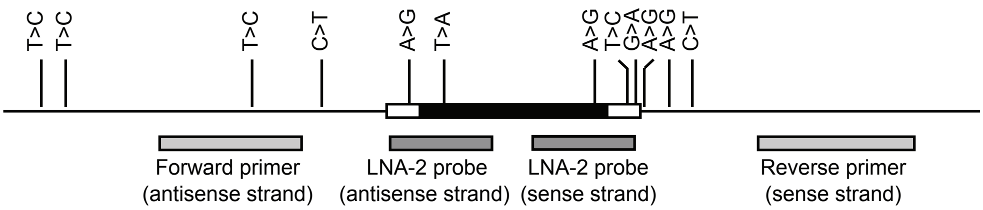


Supplementary Figure 4

A



C



B

```
                                                                                          Sal1    LoxP                                Sal1
01       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTACGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
02       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
03       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
04       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
05       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
06       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
07       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
08       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
09       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
10       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
11       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
12       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCCATACCCTGTGGACACACCCGCATGAGCTT GTCGGC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
13       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
14       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC GGCTATCCCAACACCTAC
15       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
16       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
17       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGTTAT GTCGGC AGCTATCCCAACACCTAC
18       ATGCATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
19       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAGCATACATTATACGAAGCTAT GTCGAC AGCTATCCCAACACCTAC
20       ATGTATCACGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
21       ATGTATCATGGCTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
22       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
23       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
24       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
25       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
26       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
27       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATGCGAAGTTAT GTCAAC AGCTATCCCAACACCTAC
28       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
29       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATG AGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
30       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCTAACACCTAC
31       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
32       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
33       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
34       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
35       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
36       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACATACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
37       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
38       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
39       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTATCCCAACACCTAC
40       ATGTATCATGACTTCACTCAGTACAAATATTCTGCATAGTACGCTATACCCTGTGGACACACCCGCATGAGCTT GTCGAC ATAACTTCGTATAATGTATGCTATACGAAGTTAT GTCGAC AGCTGTCCCAACACCTAC

Summary  ATGCATCACGACTTCACTCAGTACAAATATTCTGCATAGTACGCCATACCCTGTGGACATACCCGCATGAGCTT GTCGGC ATAGCTACGTATAATGTATGCTATACGAAGTTAT GTCAGC GGCTGTCCTAACACCTAC
```

Supplementary Figure 5