

# Extrachromosomal circular DNA harboring targeted tumor suppressor gene mutations supports intratumor heterogeneity in mouse liver cancer induced by multiplexed CRISPR/Cas9

**Tao Guo**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Guoqiao Chen**

Zhejiang University

**Xufan Li**

Zhejiang University <https://orcid.org/0000-0002-0775-3510>

**Meng Wang**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Xiao-Ying Yang**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Kun-Ming Liu**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Si-Cheng Liu**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Yi-Li Feng**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine

**Pengyuan Liu**

Zhejiang University

**Hui Lin**

Zhejiang University

**An-Yong Xie (✉ [anyongxie@zju.edu.cn](mailto:anyongxie@zju.edu.cn))**

Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine <https://orcid.org/0000-0002-6608-2550>

## Article

**Keywords:** CRISPR/Cas9, somatic multiplex-mutagenesis, liver tumor, intratumor genetic heterogeneity, eccDNA

**Posted Date:** February 10th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1333557/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Primary liver cancer has significant intratumor genetic heterogeneity (IGH), which drives cancer evolution and prevents effective cancer treatment. CRISPR/Cas9-induced mouse liver cancer models can be used to elucidate how IGH is developed. However, as CRISPR/Cas9 could induce chromothripsis and extrachromosomal DNA in cells in addition to targeted mutations, we wondered whether this effect contributes to the development of IGH in CRISPR/Cas9-induced mouse liver cancer. Here, using this approach targeting 34 tumor suppressor genes (TSGs), we efficiently induced autochthonous liver tumors in mice within 30-60 days. Analyses of CRISPR/Cas9-induced tumors and single-cell clones derived from tumor nodules revealed multiplexed and heterogeneous mutations at target sites. Many target sites frequently displayed more than two types of allelic variations with varying frequencies in single-cell clones, indicating increased copy number of these target sites. The types and frequencies of targeted TSG mutations continued to change at some target sites between single-cell clones and their subclones. Even the proliferation of a subclone in cell culture and in mouse subcutaneous graft altered the types and frequencies of targeted TSG mutations in the absence of continuing CRISPR/Cas9 genome editing, indicating a new source outside primary chromosomes for the development of intratumor genetic heterogeneity (IGH) in these liver tumors. Karyotyping of tumor cells revealed genomic instability in these cells manifested by high levels of micronuclei and chromosomal aberrations including chromosomal fragments and chromosomal breaks. Sequencing analysis further demonstrated the generation of extrachromosomal circular DNA (eccDNA) harboring targeted TSG mutations in these tumor cells. These findings suggest an important contribution of eccDNA carrying TSG mutations to tumor evolution in mouse liver cancer induced by multiplexed CRISPR/Cas9.

# Introduction

Cancer is driven by genetic changes, which include point mutations, small insertions and deletions (indel), large genome rearrangements and copy number alterations (CNAs)<sup>1</sup>. These genetic alterations in a single genome manifest genomic instability, a hallmark of cancer, and are generated directly or indirectly by repair of DNA damage induced by exogenous agents or from endogenous sources. This mutational process continues to operate throughout tumor evolution and may lead to phenotypic variations between cells within the final neoplasm. The genetic and phenotypic heterogeneity within a tumor nodule, defined as intratumor heterogeneity (ITH), fuels tumor growth that would follow the fundamental principles of Darwinian evolution<sup>2,3</sup>. In this model, genetic mutations are continuously generated and gradually selected for differential fitness of daughter cells in tumor growth. However, growing evidence indicates that diverse genetic alterations may occur early in bursts in many types of cancers and their pervasive distributions in tumors often fit neutral growth pattern<sup>2,4</sup>. Catastrophic events such as chromothripsis could also initiate transformation of tumor cells and result in ITH early in one or a few cell divisions of tumor cells<sup>4-8</sup>. Thus, punctuated evolution or the “Big Bang” theory associated with these mutational processes has been proposed to explain the observation of more extensive ITH than that predicted by positive selection in many types of cancers<sup>8-14</sup>.

Molecular analysis of primary liver cancer revealed significant ITH, which is attributable to the difficulty in effective treatment of liver cancer<sup>15</sup>. Primary liver cancer ranks the seventh in term of cancer incidence and the third in cancer-related mortality worldwide. Hepatocellular carcinoma (HCC) is the most common form of liver cancer and accounts for ~90% of cases<sup>16</sup>. Despite improved understanding of liver cancer, the ITH development remains one of the fundamental questions in biology of this disease. Mouse models provide an important tool to study the ITH of liver cancer<sup>17</sup>. CRISPR/Cas9, a powerful gene editing technology, has been used to rapidly generate mouse models for human cancer by inactivating tumor suppressor genes (TSGs) or activating oncogenes in somatic cells<sup>18</sup>. CRISPR/Cas9-mediated targeted inactivation of *p53* and *Pten* in mouse liver induced intrahepatic cholangiocarcinoma (ICC), a second form of primary liver cancer, in mice<sup>19</sup>. Multiplexing CRISPR/Cas9 has also been explored for high-throughput analysis of TSGs for initiation and development of liver cancer<sup>20-23</sup>. In these applications, CRISPR/Cas9 induces DNA double strand breaks (DSBs) in somatic cells and repair of these DSBs by either non-homologous end joining (NHEJ) or homology-directed repair (HDR) generates edited products that induce neoplastic transformation<sup>24</sup>.

Recent studies indicate that CRISPR/Cas9-induced DSBs could also lead to chromothripsis and extrachromosomal DNA (ecDNA) in cells, in addition to targeted mutations<sup>25,26</sup>. EcDNA generated from chromosomal DNA are often circularized to form extrachromosomal circular DNA (eccDNA), and can be inherited and integrated back into the genome of tumor cells during tumor growth, thus expanding intratumor genetic heterogeneity (IGH)<sup>27-29</sup>. It was recently reported that oncogenes on eccDNA with copy number oscillation drive tumor evolution and resistance to cancer therapy, although the contributions of eccDNA remains difficult to identify in the cancer genome due to relatively low incidence and technological barriers<sup>28,30-32</sup>. In CRISPR/Cas9-based development of mouse models for primary liver cancer, simultaneous induction of numerous DSBs by multiplexing CRISPR/Cas9 may amplify genomic instability that is potentially marked with eccDNA, adding a layer of genetic variations into ITH of mouse liver cancer<sup>25,26</sup>. Additionally, many targeted mutations of TSGs in the founder tumor cells may not be a driver, but they could support the development of ITH and participate tumor evolution in the later stage of tumor development. It is however unclear how these targeted mutations, a driver or not, engage the ITH development. Here, by simultaneously targeting 34 TSGs, we rapidly generated CRISPR/Cas9-based liver cancer in mice. By tracking targeted mutations during clonal expansion of single primary liver tumor cells derived from mouse liver tumor nodules, we discovered that the copy number of targeted mutations oscillated over time, even between null and not null, in vitro and in vivo. We found that these cells were associated with significant genomic instability and contained eccDNA harboring targeted TSG mutations. This suggests that eccDNA carrying TSG mutations may support the development of ITH in primary liver cancer.

## Results

### Induction of primary mouse liver cancer by multiplexed CRISPR/Cas9 targeting 34 TSGs

To induce mouse liver tumors using multiplexed CRISPR/Cas9, we chose 34 TSGs along with a 5'-upstream site of the *Setd5* locus as a negative control for CRISPR/Cas9-mediated gene inactivation (**Fig. 1A and Fig. S1**). These 34 TSGs are involved in at least 11 cancer-related signaling pathways (**Fig. 1A**). We designed and constructed at least 3 sgRNAs for each target gene except *p53*, and analyzed these sgRNAs for their efficiency of frame-shift editing using our BGN reporter established previously (**Fig. S2A**)<sup>33</sup>. The 23-bp target sequence containing the PAM for each sgRNA was inserted into the I-SceI-EcoRI site of the BGN reporter (**Fig. S2A**). To test each sgRNA, the expression plasmids for Cas9 and sgRNA and the BGN reporter plasmid containing the Cas9-sgRNA target were transfected into mouse embryonic fibroblast cell line NIH-3T3 cells and the frequencies of “3n+1”-bp frame-shift measured by flow cytometry for GFP<sup>+</sup> cells (**Fig. S2B**). After testing of each sgRNA, we selected the most effective sgRNA for each target gene to establish a plasmid library of 35 sgRNAs. We then used hydrodynamic tail vein injection (HTVI) to deliver this sgRNA plasmid library mixed with the *Streptococcus pyogenes* Cas9 (*SpCas9*) expression plasmid into mouse liver cells to induce liver tumors<sup>19,34</sup>. Total volume injected was 2mL with the amount of the *SpCas9* plasmid fixed at 200 µg for each mouse. The amounts of each sgRNA in the sgRNA library injected ranged from 0.005 µg to 5 µg (**Fig. 1B**). Liver tumors were formed and visible in mice within 30-60 days (**Fig. 1B-C**). Dilution of sgRNA delayed development of Cas9/sgRNA-induced mouse liver cancer and reduced cancer induction (**Fig. 1B**). After each sgRNA was reduced to 0.2 µg, tumor development rarely occurred (**Fig. 1B**). Histological analysis of tumor nodules indicated that the hepatic lobule structures were destructed in tumor nodules (**Fig. 1D**). Staining for the HCC and ICC biomarkers alpha-fetoprotein (AFP), cytokeratin 19 (CK19) and Golgi glycoprotein 73 (GP73) indicated induction of HCC, ICC and mixed HCC-ICC type in CRISPR/Cas9-induced mouse liver tumors (**Fig. 1D**). In one section of a single tumor nodule, three selected regions exhibited distinct histologic features (**Fig. 1E**), indicating strong ITH in CRISPR/Cas9-induced liver cancer in mice.

### Heterogeneity of targeted TSG mutations in CRISPR/Cas9-induced mouse liver tumors

We performed next generation sequencing (NGS) of PCR-amplified target sites in tumor nodules to identify targeted TSG mutations and determine the frequencies and spectra of these mutations. We define an allelic mutation with a frequency at no less than 5% as a true mutation to reduce the interference of sequencing errors. While different targets showed different mutational spectra, we also found that many same target sites carried different allelic mutations with varying frequencies. For example, in one analysis, in addition to their respective wild-type (WT) allele, the *p53* target site had 3 types of mutations, 28.12% for 1-bp deletion on the left side of the break (termed Del1|0), 16.13% for 7-bp deletion on the right side of the break (Del0|7) and 13.87% for insertion of 1A at the break (Ins1A); *Atm* had two (23.81% for D6|0 and 20.07% for Ins1C) and *Rb1* had none in addition to WT alleles (**Fig. 2A**). In each tumor, majority of target genes were mutated at intended sites and the mutation frequencies (MF) were high at some of these sites after the frequencies of all allelic mutations were combined for each site (**Fig. 2B-C, and Table S1**). The mutation profiles differed between tumors within a same mouse or from

different mice, indicating intertumor heterogeneity in CRISPR/Cas9-induced liver cancer with respect to targeted mutations of TSGs.

In addition, a single tumor nodule was microdissected into four regions, and each region analyzed for targeted TSG mutations by NGS of PCR-amplified target sites. Significant variations existed in mutation profiles between these four regions (**Fig. 2C and Table S1**), suggesting ITH of targeted TSG mutations. These variations may be also attributable to varying healthy stromal components and/or residual normal tissue in individual tumor nodules.

To further analyze ITH of targeted TSG mutations, we established single-cell clones from tumor nodules and analyzed targeted mutations in these tumor cells by targeted PCR amplicon deep sequencing. While about half of 34 TSGs were highly mutated between individual clones from the same nodules (in red in **Fig. 2C**), the frequencies of targeted mutations varied widely between 0% and 100% in the remaining genes (from *Smad4* to *Hnf1a*; **Fig. 2C**). Between different clones from the same nodules, some targets sites carried identical mutations with similar frequencies but some harbored identical mutations with different frequencies or even different mutations (**Fig. 2C and Table S2**). These results indicated strong ITH of targeted TSG mutations in CRISPR/Cas9-induced mouse liver cancer.

As observed previously<sup>22</sup>, individual target sites often carried more than two different allelic mutations within a tumor nodule (**Table S1**). This could be explained at least by the following three possibilities: 1) Several transfected founder cells happened to start together with a different subset of mutations induced by multiplexed CRISPR/Cas9 genome editing and developed into a single tumor nodule. 2) Some mutations were first induced in a single transfected founder cell by multiplexed CRISPR/Cas9 genome editing and others occurred only after the first cell division in subsequent daughter or granddaughter cells<sup>22</sup>. 3) More than two copies of target sites exist in the genome of a single transfected founder cell and are differently mutated by CRISPR/Cas9. However, in some single-cell clones (e.g., 1C3 and 6C7), a few target sites harbored more than two different allelic mutations and even up to 9 different allelic mutations, some of which occurred with varying frequencies (**Table S2**). This was unexpected for a single-cell clone where only two copies of a target site are the most likely for CRISPR/Cas9-induced mutations unless the cells are hyperploid or the target sites have repeat sequences. Nevertheless, this type of copy number variations (CNVs) at the single-cell level represents a new source for heterogeneity of targeted TSG mutations in CRISPR/Cas9-induced mouse liver cancer.

### **Type and frequency of targeted mutation alterations from parental clones to subclones**

Given the presence of more than two allelic mutations with significant different frequencies in a single-cell clone, we wondered whether this mutation pattern is inheritable. We thus selected two single-cell clones (i.e., 1C3 and 6C7) that carried more than two copies of allelic mutations at some target sites, isolated 4 subclones from 1C3 and 8 subclones from 6C7, and analyzed targeted mutations in these clones and subclones. The profile of target site mutations was different between parental clones and

their respective subclones and between subclones (**Fig. 3A and Table S3**). For instance, 6 *Rb1* target site variants, 8 *Lkb1* target site variants, 4 *Arid1a* target site variants and 5 *Smad4* target site variants were detected across parental 1C3 clone and its 4 subclones, but each type displayed different frequencies within individual clones and the mutation profiles were also different between clones (**Fig. 3B-E**). The frequency of the wide type *Rb1* was less than 2% in paternal clone 1C3, but over 30% in 4 subclones (**Fig. 3B**). In contrast, the frequency of the *Rb1* Del22|5 mutation was nearly 30% in 1C3, but hardly detectable in subclone 1C3-1 or 1C3-2, less than 3% in 1C3-3 and about 5% in 1C3-4 (**Fig. 3B**). The *Lkb1* Del4|0 mutation was barely detected in 1C3, but the frequencies of this mutation were 20% or more in 4 subclones (**Fig. 3C**). The *Lkb1* Del10|0 mutation was a dominant mutation with the frequency at over 30% in 1C3 and its subclone 1C3-4, but less than 1% in subclone 1C3-1 and 1C3-2 and around 16% in subclone 1C3-3 (**Fig. 3C**). The *Arid1a* Del11|16 mutation was dominant at more than 70% in 1C3-1 but negligible at less than 2% in 1C3-4 (**Fig. 3D**). In contrast to Del11|16, the *Arid1a* Del10|9 mutation was nearly undetectable in 1C3-1 but highly frequent at more than 50% in 1C3-4 (**Fig. 3D**). Similarly, while the *Smad4* WT allele was detected at over 60% in 1C3 but less than 6% in 1C3-4, the frequency of the *Smad4* Del2|0 was 2% in 1C3 but nearly 35% in 1C3-4 (**Fig. 3E**). These results indicate continuing oscillation in the frequencies of target site mutations between these parental clones and subclones. In particular, predominant mutations in parental clones could disappear in subclones whereas negligible mutations in parental clones could appear in subclones.

As spontaneous mutation at a given site normally occurs with extremely low probability during cell proliferation<sup>35</sup>, it is surprising that the frequency of a specific targeted mutation oscillates significantly from parental single-cell clones to daughter single-cell subclones. We speculated that Cas9-sgRNA might be stably expressed in clones isolated and continue to edit the WT target sites, thus altering the frequency of the targeted mutation. However, we did not detect *SpCas9* proteins in the parental clones 1C3 and 6C7 and their subclones (**Fig. S3A**). It remains possible that a small amount of *SpCas9* proteins stably synthesized in the cells could actively mutate their target sites, even if the protein level is hardly detectable by Western blot. We thus transfected these parental clones or subclones with the expression plasmids for two *Col1a1* sgRNAs and one *Rosa26* sgRNA with or without the *SpCas9* expression plasmid and measured the editing frequencies at the target sites by PCR amplicon deep sequencing. While the editing is efficient with transfection of both *SpCas9* and sgRNA, no editing was detected with sgRNA transfection alone or with neither *SpCas9* nor sgRNA (**Fig. S3B-C**). This excludes the possibility that continuing CRISPR/Cas9 genome editing alters the types and frequencies of targeted TSG mutations from parental clones to subclones.

### Continuing type and frequency oscillation of targeted mutations during clonal expansion

Next, we asked whether the types and frequencies of targeted mutations change over time during expansion of a subclone. We continuously cultured 4 subclones (i.e., 1C3-1, 1C3-2, 1C3-3 and 1C3-4) and analyzed target site mutations of each subclone on Day 0, Day 15 and Day 30 (**Fig. 4A and Fig. S4**). The

frequencies of some targeted mutation changed over time during clonal expansion, e.g., from 2.47% at Day 0 to 35.09% at Day 15 and 34.09% at Day 30 for *Rb1* Ins1T in 1C3-1 (**Fig. S4 and Table S4**). After storage in liquid nitrogen for a year, we thawed and continued to culture the 1C3-1 clone. We repeated analysis of target site mutations in the cultured cells on day 0 (i.e., Mon12), day 90 (i.e., Mon15) and day 180 (i.e., Mon18) (**Fig. 4A**). The frequencies of some targeted mutations continued to change over time during cell culturing of 1C3-1 (**Fig. 4B and Table S4**). For example, the *Rb1* Ins1T mutation was negligible on Day 0 but appeared with the frequencies at over 30% on Day 15 and Day 30. Del1|0 mutation was significant at over 60% on Day 0 and decreased to about 30% on Day 15, Day 30 and Mon 12 and even 10% on Mon 15. On Mon 18, this mutation was slightly increased to over 20% (**Fig. 4C**). Both *Lkb1* Del5|2 and Ins1G mutations remained frequent at over 40% from Day 0 to Mon 12, but decreased to about 20% on Mon 15 and increased again to 30% on Mon 18 (**Fig. 4D**). *Arid1a* Del11|16 and Del10|9 mutations changed in opposite direction on Mon 15 and Mon 18 (**Fig. 4E**). Similarly, *Smad4* WT and Del16|0 mutations oscillated in opposite direction on Mon 15 and Mon 18 (**Fig. 4F**). This again confirms that some targeted mutations are not stable during proliferation of single tumor cells derived from CRISPR/Cas9-induced liver cancer in mice.

### Type and frequency alteration of targeted mutations from single-cell clones to subcutaneous grafts

In order to evaluate this instability of targeted TSG mutations in vivo, we implanted the subclone 1C3-1 cells into 4 immunodeficient SCID mice subcutaneously (**Fig. 5A**). All mice grew a visible tumor in about 5 days. We harvested the tumor tissues from these 4 mice at the 14<sup>th</sup> day and established 2 single tumor cell clones from each tumor tissue (**Fig. 5A**). Analysis of targeted mutations revealed that tumor tissues harbored more frequent WT allele than both 1C3-1 and tumor cell clones derived from tumor tissues, likely due to the presence of normal cells in subcutaneous tumor samples (**Table S5**). Therefore, in order to exclude the interference of normal cells from tumor tissues on the frequency of targeted mutations in tumor cells, we compared the frequencies of targeted mutations only between parental 1C3-1 clone and tumor cell clones (**Fig. 5B**). The frequencies of some targeted mutations changed significantly among these tumor cells (**Fig. 5B and Table S5**). For instance, 1bp deletion (Del1|0) of *Rb1* was dominant with the frequencies at over 50% in SG1-1, SG1-2, SG2-5 and SG2-6, but infrequent at less than 5% in SG4-1, SG4-3, SG5-1 and SG5-2 (**Fig. 5C**). Differently, the parental subclone 1C3-1 harbored this *Rb1* Del1|0 mutation at 40% (**Fig. 5C**). Furthermore, while the Del5|2 mutation of *Lkb1* changed little with the frequency at about 40% among all tumor cells tested, the *Lkb1* Ins1G mutation oscillated significantly among these tumor cells, with the frequency at about 40% in SG1-1, SG1-2, SG2-5, SG2-6 as well as 1C3-1, but nearly undetectable in SG4-1, SG4-3, SG5-1 and SG5-2 (**Fig. 5D**). In addition, two *Col1a1* mutations, i.e., Del1|0 and Ins1T, oscillated in an opposite direction (**Fig. 5E**). While the *Col1a1* Del1|0 mutation dominant in 1C3-1, SG1-1, SG1-2, SG2-5 and SG2-6 was negligible in SG4-1, SG4-3, SG5-1 and SG5-2, the *Col1a1* Ins1T mutation infrequent in 1C3-1, SG1-1, SG1-2, SG2-5 and SG2-6 occurred frequently in SG4-1, SG4-3, SG5-1 and SG5-2 (**Fig. 5E**). Similarly, the Del1|0 and Ins1A mutations of *Rasa1* started with the frequencies at around 50% in parental 1C3-1 clone and then oscillated in a reverse pattern among single-

cell clones from tumor grafts (**Fig. 5F**). These results indicate that targeted mutations in tumor cells derived from CRISPR/Cas9-induced liver cancer in mice are also unstable in subcutaneous grafts derived from tumor cells with the frequencies of some targeted mutations oscillating in vivo. Of note, the WT alleles of *Rb1*, *LKb1*, *Col1a1* and *Rasa1* not shown in **Fig. 5B** were all below 5% in frequency among the clones analyzed, but indicated in **Fig. 5C-F** as a negative comparison.

### Increased genomic instability in CRISPR/Cas9-induced mouse liver cancer cells

As mentioned previously<sup>35</sup>, spontaneous mutations at a given site normally occurs at an extremely low rate and are impossible to cause significant alterations in the types and frequencies of targeted mutations detected in the study. Continuing targeted editing by Cas9-sgRNA was also excluded as a causal factor because neither stable expression nor the editing activity of Cas9-sgRNA was detected in these single-cell clones. As cancer with genomic instability has an increased tendency for constant genomic alteration, we wondered whether these primary liver tumor cells are associated with strong genomic instability. We first examined spontaneous  $\gamma$ -H2AX and 53BP1 foci formation in 10 primary tumor cell lines derived from CRISPR/Cas9-induced mouse liver cancer, NIH3T3 and the mouse liver cancer cell line Hepa1-6 (**Fig. 6A**). All of these liver cancer cell lines except 1C3-2 showed a higher level of spontaneous  $\gamma$ H2AX focus formation than NIH3T3 (**Fig. 6A-B**), indicating strong induction of spontaneous DNA DSBs and activation of DNA damage response in these cancer cells. It was however unclear why 53BP1 focus formation was much less frequent than  $\gamma$ H2AX focus formation in nearly all of these cell lines as both  $\gamma$ H2AX and 53BP1 foci indicated the site of DSBs (**Fig. 6A-B**).

We also analyzed these CRISPR/Cas9-induced primary liver tumor cells for micronucleus formation, which is frequently involved in chromosomal aberrations and genomic instability in cancer<sup>36</sup>. Micronucleus formation was readily detected in these tumor cells (**Fig. 6C**). The percentages of micronucleated cells were significantly higher in tumor cells than in NIH3T3 cells (**Fig. 6D**). In particular, while about 1% of NIH3T3 cells contained micronuclei, over 10% of 1C3 or 1C3-1 cells were micronucleated (**Fig. 6D**). Metaphase spread analysis further identified significant chromosomal aberrations in primary liver tumor cells (**Fig. 6E**). The number of chromosomes in primary liver tumor cells varied in average from 57.4 in 1C3 to 105.7 in 6C7-2 and was much more than 40 in a diploid mouse cell (**Fig. 6E-F**). This indicates that these tumor cells are hyperploid, at least in part contributing to more than two allelic variations at a target site of TSGs in single-cell clones. Biarmed chromosomes appeared in all Hepa1-6 cells as reported previously<sup>37</sup>, but only existed in 2.0-6.0% of the primary liver tumor cell lines 1C3, all 1C3 subclones and 6C7-5 as well as in 8.9% of NIH-3T3 (**Fig. 6E-F**). In contrast, we did not detect any biarmed chromosomes in 6C7 or its subclones 6C7-2, 6C7-3 and 6C7-4. The difference in the generation of biarmed chromosomes between 6C7-5 and its parental clone 6C7 or other 6C7 subclones indicates potential variations in genomic instability in these cells (**Fig. 6E-F**). In addition, primary liver tumor cells exhibited more frequent chromosomal fragments than NIH-3T3 and Hepa1-6 (**Fig. 6E-F**). As frequent micronucleus formation, chromosomal aberrations and chromosomal fragments in primary

tumor cells manifested the genomic instability of CRISPR/Cas9-induced mouse liver tumors, these findings indicated a possible connection between the genomic instability and the oscillation in the frequencies of targeted TSG mutations in tumor cells. Variations in genomic instability among parental clones (i.e., 1C3 and 6C7) and their respective subclones again implied that an intrinsic genetic force drive the oscillation of target site mutations in these single-cell clones.

### Detectable eccDNA harboring targeted site mutations

Due to unstable nature of chromosomal fragments, the frequent presence of small chromosomal fragments observed in our study could drive significant alterations in the frequencies of targeted mutations in tumor cells derived from CRISPR/Cas9-induced mouse liver cancer. However, because it was difficult to separate chromosomal fragments from the regular chromosomes, we could not determine the contribution of chromosomal fragments to the oscillation in the frequencies of targeted mutations in tumor cells. Instead, like chromosomal fragments, eccDNA could be formed in CRISPR/Cas9-induced mouse liver cancer; this form of ecDNA could however be separated from genomic DNA (gDNA) and analyzed for targeted mutations by targeted PCR amplicon deep sequencing (**Fig. S5A**). Indeed, using eccDNA isolated from 1C3-1, 1C3-2, 1C3-3, 6C7-2 and 6C7-4 as the template for targeted PCR, we found that many of 35 target sites were poorly amplified as compared to gDNA and the numbers of deep sequencing reads for these sites were low, indicating the absence of eccDNA harboring these sites (**Fig. S5B, and Table S6**). Only 13 targets (i.e., *Atm*, *Cdkn2a*, *Col1a1*, *Fbxw7*, *Kmt2c*, *Nf1*, *Notch3*, *p53*, *Pik3r1*, *Rb1*, *Rps6ka3*, *Setd5* and *Tet2*) consistently exhibited sufficient reads in eccDNAs (**Fig. S5B and Fig. S6**). Deep sequencing of targeted amplicons further confirmed that eccDNA in 1C3-1, 1C3-2, 1C3-3, 6C7-2 and 6C7-4 clones contain intact TSG target sites and targeted TSG mutations (**Table S4 and Fig. S6**).

To compare the distribution of targeted mutations across 34 target sites between eccDNA and gDNA, we calculated percentages of combined reads from all 34 target sites for reads of each target site respectively in both eccDNA and gDNA and derived relative read ratio of eccDNA to gDNA for each target site in 1C3-1 by dividing percentage values for eccDNA to respective ones for gDNA (**Fig. S5C-D**). Because PCR variations of eccDNA were normalized by PCR variations of gDNA in this relative read ratio of eccDNA to gDNA, the interference of PCR variations on sequencing reads in eccDNA across target sites were partly excluded. While the profiles of targeted mutations were different between eccDNA and gDNA as expected, it was surprising that some mutations undetectable or with a low frequency in gDNA increased significantly in frequency in eccDNAs (**Fig. 7A, Fig. S5C-D and Fig. S6**). For example, the frequency of *Atm* Del3|0 mutation was less than 2.35% in gDNA, but was elevated to 17.88% in eccDNA (**Fig. 7B**). Similarly, the frequency of the *Rb1* Del22|5 mutation is 9.95% in gDNA, but increased to 28.53% in eccDNA (**Fig. 7C**). In addition, some targeted mutations detected at a high rate in gDNA were undetectable or much less frequent in eccDNA harboring these targets (**Fig. 7A, Fig. S5D, Fig. S6 and Table S6**). For instance, the frequency of the *Rb1* Del1|0 mutation was 23.82% in gDNA, but was lowered to 7.53% in eccDNA (**Fig. 7C**).

To directly identify eccDNA harboring targeted mutations, we designed outward PCR primer pairs to determine circularization junctions of eccDNA carrying the respective targets whereas inward primer pairs were usually used to identify targeted mutations (**Fig. S7A**). Because 13 targets were frequently found in eccDNA isolated from 1C3-1, we followed these 13 sites to design outward PCR primer pairs and identify circularization junctions of eccDNA. Because the outward PCR primer pairs are positioned at one side to targeted sites, target sites with or without any variations could also be identified (**Fig. S7A**). Some of these 13 targets generated PCR bands with outward primer pairs, and PCR products with outward primer pairs were purified and cloned into a pUC19 vector for Sanger sequencing. We detected the circularization junctions of eccDNA harboring either intact target site or targeted mutations of *Atm*, *Rb1* and *Kmt2c* (i.e., *Atm*, *Rb1* and *Kmt2c* eccDNA) (**Fig. 7D-F and Fig. S7B**). Based on positions of circularization junctions and targeted mutations, we determined that the sizes of *Atm*, *Rb1* and *Kmt2c* eccDNA were about 200 bp, 211 bp and 213 bp, respectively (**Fig. 7D-F**). Among 10 clones sequenced for *Atm* on eccDNA, 1 contained the WT sequence at the target site and Ins1G at the circularization junction, 1 contained Del3|0 at the target site and Del1T Ins1C at the circularization junction, and 8 contained Del8|0 Ins4CCCG at the target site and no deletion/insertion at the circularization junction (**Fig. 7D**). Among 4 clones sequenced for *Rb1* on eccDNA, all had Del22|5 at the target site and were joined accurately at the circularization junction, whereas 1 clone from *Kmt2c* eccDNA had WT target site and no deletion/insertion at the circularization junction (**Fig. 7E-F**). Most of these variations such as *Atm* WT, *Atm* Del3|0, *Rb1* Del22|5 and *Kmt2c* WT were also present in gDNA (**Fig. 7B-C and Table S6**). Taken together, these results demonstrated the existence of eccDNA harboring some targeted mutations of TSGs in CRISPR/Cas9-induced mouse liver cancer. As eccDNA is unstable in varying copies and could be degraded or integrated back to regular chromosomes<sup>38</sup>, eccDNA harboring targeted mutations may provide genetic materials to alter the frequencies of targeted mutations in the genome, driving ITH in CRISPR/Cas9-induced mouse liver cancer.

## Discussion

CRISPR/Cas9-based targeted somatic gene editing of TSGs has been used to induce HCC in mice<sup>19–23</sup>, providing a useful tool to study the biology of HCC, one of the leading causes of cancer-related death worldwide. By analyzing targeted mutations of TSGs, our work describes a characterization of IGH in mouse model of liver cancer induced by multiplexed CRISPR/Cas9. Specifically, our data show: (1) Strong heterogeneity with respect to the types and frequencies of targeted TSG mutations between tumor nodules and within a tumor nodule; (2) Significant variations in allelic mutation type and frequency at a TSG target site between single-cell clones; (3) More than two types of allelic variations with varied frequencies at many of TSG target sites within a single-cell clone; (4) Oscillation of allelic variations in type and frequency at some of TSG target sites in subclones, during proliferation of a subclone and in mouse subcutaneous graft of a subclone. In addition, these tumor cells exhibited strong genomic instability manifested by micronucleation, chromosomal aberrations, hyperploidy and generation of eccDNAs. In particular, while polyploidization is a characteristic feature of hepatocytes<sup>39</sup>, hyperploid liver tumor cells and altered hyperploidy could change the copy number of genes including targeted TSGs

during cell division. In the meantime, eccDNA could cause CNVs by random integration into a chromosome during cell proliferation<sup>7,27,29</sup>. This may help explain persistent oscillation of allelic TSG mutation types and frequencies in single-cell clones. Together, this study identifies a potential source for IGH in mouse liver cancer induced by multiplexed CRISPR/Cas9.

Since being first discovered as double minutes (DMs) in 1965<sup>40</sup>, eccDNA have recently emerged as important genetic molecules, whose functions are implicated in normal cellular function and the development of human diseases including cancers<sup>27,29,38</sup>. However, how eccDNA is generated in mammalian cells remains elusive although this question has been under intensive investigation. Many mechanisms such as the breakage-fusion-bridge (BFB) cycle, micronucleation, chromothripsis and cell death-induced DNA fragmentation have been proposed to explain the biogenesis of eccDNA and may involve DNA breakage and repair<sup>41-46</sup>. The BFB cycle caused by chromosomal fusion is regarded as an initiating factor for a cascade of events such as micronucleation, chromothripsis and even cell death, thus generating eccDNA<sup>44,47</sup>. In micronuclei, chromosomes are often susceptible to DNA breakage due to incomplete nuclear envelope, leading to chromothripsis, a single catastrophic event where a chromosomal region is shattered into a number of fragments<sup>8,46</sup>. While re-integration of micronuclei into primary nuclei during mitosis could allow DNA fragments to engage complex chromosomal rearrangements in the genome, some DNA fragments are circularized to form eccDNA. Similar to eccDNA generation, reintegration of eccDNA elements to the genome may also require DNA damage and DNA recombination although the underlying mechanisms are not well understood<sup>7,29,46</sup>. EccDNA could also result from apoptotic DNA fragmentation and may enter neighboring cells to induce an innate immune response<sup>45</sup>. Given that tumor cells from mouse liver cancer model contained high levels of micronuclei and chromosomal aberrations as well as detectable eccDNAs, it is likely that chromosomal fusion, the BFB cycle and chromothripsis also occur, but this likelihood has yet to be confirmed. Our data have also shown that tumor cells are hyperploid in mouse liver cancer model induced by multiplexed CRISPR/Cas9. As chromothripsis can be promoted by hyperploidy<sup>48</sup>, this suggests an additional mechanism of eccDNA generation.

Recent studies have identified oncogene-containing eccDNA and demonstrated that oncogenes on eccDNA are commonly amplified in cancer<sup>28-32</sup>. Due to usual lack of centromeres and unequal segregation, eccDNA can also drive the development of IGH<sup>7,27,29</sup>. Such eccDNA can markedly increase oncogene copy number, promote accessible chromatin and high oncogene expression, drive IGH and accelerate tumor evolution. However, extra copies of TSGs on eccDNA and their inactivating mutations were not expected to add more cancer-promoting capacity as a driver mutation than inactivating mutations on genomic TSGs. Thus, eccDNAs harboring TSG mutations are often ignored although they may support the development of ITH in punctuated tumor evolution. It is yet to be determined whether eccDNA harboring TSG mutations are present or function in cancer cells and if so, what their functions are.

Previous study has indicated that the use of multiplexed CRISPR/Cas9 could simultaneously generate many on-target and off-target DSBs in a cell, leading to fragmented chromosomes, micronuclei formation and undesired chromosomal rearrangements including insertions/deletions, translocations, duplications and complex rearrangements<sup>22,25</sup>. Even one single DSB induced by CRISPR/Cas9 is capable of initiating the BFB cycle, consequently leading to micronucleation, chromothripsis, aneuploidy or cell death<sup>47</sup>. Chromosomal fragments generated either directly from DNA cutting by multiplexed CRISPR/Cas9 or indirectly from chromothripsis or cell death could be circularized by DNA ligases, forming eccDNA<sup>45</sup>. Even linear chromosomal fragments generated by paired CRISPR/Cas9 could form eccDNA, which can in turn be reintegrated into the genome for stable expression, in human cells<sup>26</sup>. In this study, we used multiplexed CRISPR/Cas9 to target TSGs in mouse liver and generated mouse models of liver cancer. During targeted TSG disruption in mouse liver cells, extrachromosomal fragments could arise to form eccDNA either with or without targeted TSG mutations in the transfected founder cells. EccDNA with intact TSG target sites could be targeted again by CRISPR/Cas9 to generate targeted mutations if CRISPR/Cas9 remains active in early stage of mouse liver cancer development. EccDNA carrying targeted TSG mutations or no targeted mutations could be reintegrated into the genome of proliferating cells by NHEJ or HDR and induce CNVs in daughter cancer cells, promoting IGH<sup>7</sup>. In addition, it appears that transient expression of multiplexed CRISPR/Cas9 is sufficient to initiate persistent oscillation of allelic TSG mutation types and frequencies in single-cell clones as well as strong genomic instability, as a single DSB can induce iterations of chromothripsis in tumor cells<sup>47</sup>. This indicates that persistent generation of eccDNA in CRISPR/Cas9-induced liver cancer may be assisted by the strong genomic instability manifested by highly activated endogenous DNA damage response, micronuclei formation, increased chromosomal aberrations and hyperploidization in tumor cells.

Derived from gDNA naturally, eccDNA range in size from a few hundred bases to megabases, but majority of eccDNA are smaller than 1 kb<sup>28,38,49,50</sup>. In this study, we have identified a few small eccDNA with or without targeted TSG mutations from single-cell clones of mouse liver cancer model induced by multiplexed CRISPR/Cas9. The size of these eccDNA is around 200 bp and could not carry a full TSG gene but only TSG fragments with targeted mutations or no targeted mutations. These TSG fragments on their own may not function but exert their effect upon integration into the chromosomes. For example, the TSG fragments on eccDNA could replace allelic region of TSGs in the chromosomes of some cells by HDR and alter mutation types in these cells. Thus, the types of targeted mutations and the frequencies of each TSG mutation type oscillate during proliferation of single-cell clones. In some cases, if eccDNA harboring targeted TSG mutations are present, active WT TSGs in tumor cells could be inactivated by HDR with these eccDNA as a template during the development of IGH. Additionally, TSG fragments carrying targeted mutations on eccDNA could be randomly inserted into the chromosomes by NHEJ, not only destabilizing the genome but also potentially initiating a mutational cascade for TSG mutations and CNVs. Therefore, like any other small eccDNA, small eccDNA carrying TSGs or TSG fragments with or without mutations could not only be generated but also serve as a fuel for continuing genetic variations in chromosomal TSGs during proliferation of cancer cells, driving IGH and tumor evolution in mouse models of liver cancer induced by multiplexed CRISPR/Cas9. This also suggests a possibility that

eccDNA carrying TSGs or TSG mutations may be among the sources for the development of IGH in human cancer.

As oncogene-containing eccDNA promotes drug resistance of tumor cells<sup>28,29,51</sup>, this prompted us to test whether eccDNAs harboring targeted TSG mutations affect the development of drug resistance in CRISPR/Cas9-induced mouse liver cancer. However, after treating mouse liver tumor cells that contain eccDNA carrying targeted TSG mutations with several drugs including oxaliplatin, olaparib and sorafenib, we have not yet been able to identify any definitive relationship between drug treatment and oscillation of targeted TSG mutations (data not shown). It is possible that disruption of multiple signaling pathways in these tumor cells may limit the effect of eccDNA with targeted TSG mutations on the sensitivity to drug and the development of drug resistance. This application could be improved by further reduction of targeted TSG number in multiplexed CRISPR/Cas9-induced mouse liver cancer. Nevertheless, this study generates a mouse model that can be used to analyze IGH in liver cancer by tracking targeted TSG mutations and to explore the roles of eccDNAs harboring targeted TSG mutations in tumor evolution. In particular, single-cell clones derived from this mouse model of live cancer carry mutations at up to 33 TSGs, are resistant to drugs and could allow efficient immunocompetent allograft, thus providing liver tumor cell lines for studying the interaction between aggressive cancer cells and immune microenvironment. In addition, this study is the first to have identified eccDNAs carrying target site mutations of TSGs as a potential source for the development of IGH in mouse liver cancer induced by multiplexed CRISPR/Cas9. Thus, the relevance to human cancer, particularly human liver cancer, may thus warrant further investigation.

## Methods And Materials

### Cell culture and transfection

Murine fibroblast cell line NIH-3T3, hepatoma cell line Hepa1-6 and primary murine liver cancer cells were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum (FBS) (Gibco), 100 IU/ml penicillin (Gibco) and 100 µg/ml streptomycin (Gibco) at 37 °C with 5% CO<sub>2</sub>. Transfection in cell line was performed with Lipofectamine 2000 transfection reagent (Invitrogen) in 24-well plates according to the manufacturer's protocol.

### Target TSG selection and sgRNA design and cloning

34 target TSGs were selected according to the following criteria: 1) TSGs frequently mutated in primary human liver cancer; 2) TSGs recently reported as candidate tumor suppressor genes in primary human liver cancer. Selected TSGs were also involved in at least 11 cancer-related signaling pathways. The 5'-upstream untranslated site of the *Setd5* locus close to the safe-harbor *Rosa26* locus was chosen as a genome editing control. The *Streptococcus pyogenes* Cas9 (*SpCas9*)-sgRNA plasmid px330 was reconstructed to generate separate expression plasmids for *SpCas9* and sgRNA. Single *p53* sgRNA

sequence described previously was chosen and constructed into the sgRNA expression plasmid<sup>19</sup>. Single *Setd5* sgRNA sequence was also chosen as the genome editing control. For the remaining 33 TSGs, 3 sgRNAs each gene were designed and constructed as described previously<sup>33</sup>. All constructs were validated by Sanger sequencing.

### **Identification of suitable sgRNAs for somatic mutagenesis of 34 TSGs**

To identify a single efficient sgRNAs from 3 sgRNA candidates for each of 34 TSGs, we modified the BGN reporter previously established to determine sgRNA efficiency<sup>33</sup>. Briefly, 23-nt oligodeoxynucleotides (ODNs) for sgRNA targets were ordered, annealed and inserted into the I-SceI-EcoRI site within the linker region of the *Blasticidin S deaminase (BsdR)-GFP* fusion in the BGN reporter. As a mutagenic NHEJ reporter, the modified *GFP*-based BGN reporter transfected into cells produced out-of-frame GFP after transcription and translation and thereby no GFP<sup>+</sup> cells<sup>33</sup>. A site-specific DSB was induced at the BGN reporter in NIH-3T3 cells after co-transfection of the expression plasmids for *SpCas9* and sgRNAs with the modified BGN reporter containing the *SpCas9*-sgRNA target. Mutagenic NHEJ of this DSB could generate indels inducing “3n+1”-bp frame-shift, thus reframing GFP to the correct frame and making the cell GFP<sup>+</sup>. GFP<sup>+</sup> cells could be measured by flow cytometry. Therefore, the percentages of GFP<sup>+</sup> cells at 72 h post-transfection could reflect the efficiency of sgRNA-guided *SpCas9* editing, and effective sgRNAs be identified.

### **Animal experiments and animal work statement**

6-week male wide-type C57BL/6 mice and severe combined immunodeficient (SCID) mice were purchased from Shanghai SLAC Laboratory Animal Co., and were maintained at the Animal Research Center of the Institute of Translation Medicine School of Medicine, Zhejiang University. To induce primary mouse liver cancer, a total of 200 µg *SpCas9* plasmid and 35 µg of sgRNAs were mixed in 0.9% NaCl solution with 1.5-2.0 mL volume and the mixture that accounts for 8-10% of body weight in amount was injected into C57BL/6 mice via tail vein within 10 seconds. For dosage experiments, each sgRNA was used at 0.05-1.0 µg and the sgRNA empty vector was added to 35 µg in total for the sgRNA expression plasmids. At day 30, 90 and 120 after injection, mice were sacrificed and examined for tumor formation. Mouse livers and liver tumors were collected to make formalin-fixed paraffin-embedded (FFPE) tissues or to be fixed with liquid nitrogen for gDNA extraction. For subcutaneous tumor grafts, 1x10<sup>6</sup> 1C3-1 cells were suspended in 200 µL Phosphate Buffered Saline (PBS) and inoculated subcutaneously at inguinal region of SCID mice. The tumor formation was examined daily and the tumors harvested at day 14 post inoculation. All animal experiments were carried out according to the animal ethics guidelines and approved protocols from the Animal Care and Use Committee of Zhejiang University.

## **Mouse liver tissue dissection, histology and immunohistochemistry**

5µm slices of non-tumor and tumor FFPE tissues were stained with hematoxylin and eosin (HE) and analyzed under a microscope (Leica DM4000, Germany). Tumor tissues were evaluated for expression of AFP, CK19 and GP73 by immunohistochemistry. Briefly, 5 µm FFPE tumor sections were blocked with goat serum for 30 min, incubated with mouse anti-AFP (1:50), anti-CK19 (1:50) or anti-GP73 (1:100) monoclonal antibodies (Santa Cruz Biotechnology, USA) at 4°C overnight, and further stained with 3,3'-diaminobenzidine at room temperature for 3 min. Expression of AFP, CK19 and GP73 were examined by microscope.

## **Establishment of primary mouse liver cancer cell lines and single-cell clones**

Fresh mouse primary liver tumor tissues were obtained within 1h after mice were sacrificed under sterile conditions. The tissues were minced into less than 1 mm<sup>3</sup> small pieces using scissors, washed with PBS and centrifuged at 1000 rpm for 5 min. The tissue precipitates were digested with 0.1% trypsin at 37°C for 30 min before DMEM medium containing 10% FBS was added to stop digestion. The suspension was filtered by 400-mesh cell sieve and centrifuged at 1000rpm for 5 min. Cell pellets were resuspended in complete culture medium and plated into a 10 cm dish to establish primary mouse liver cancer cell lines. To establish single-cell clones, 100 single cells were diluted into 8mL culture medium and plated into a 10 cm dish. After 10-14 days, single clone was visible, picked under a microscope using plastic clonal ring and transferred into a 96-well plate for cell expansion.

## **Immunofluorescence for liver cancer biomarkers and DNA damage response analysis**

To analyze spontaneous γH2AX and 53BP1 focus formation in primary liver cancer cells, cells were fixed in 4% paraformaldehyde for 10 min on ice, permeabilized with 0.5% Triton 100 for 10 min on ice, and blocked with 1% bovine serum albumin for 30 min. Subsequently, cells were stained with primary antibodies mouse anti-γH2AX monoclonal antibody (Abcam) or rabbit anti-53BP1 polyclonal antibody (Abcam) overnight and then with secondary antibody for 2 h. Cell nuclei were stained by 4',6-diamidino-2-phenylindole (DAPI). γH2AX and 53BP1 foci were determined and counted under fluorescence microscope. Micronuclei were also stained with DAPI and counted. All experiments were repeated at least three times and images were captured by microscope.

## **Detection of stable Cas9 activity in cancer cell lines**

Western blotting was performed to detect *SpCas9* protein in primary liver tumor cell lines derived from CRISPR/Cas9-induced mouse liver tumors. Cells were washed with cold PBS for two times, and then

lysed in Radio-Immunoprecipitation Assay buffer containing 1x protease inhibitor and 1mM PMSF on ice for 15min. After centrifugation of lysis extracts at 4°C, 10 µg proteins from suspension were separated on 10% sodium dodecyl sulfate polyacrylamide gel and transferred onto polyvinylidene fluoride membranes. Membranes were blocked with non-fat milk at room temperature for 1h, washed with TBST buffer, and incubated with mouse anti-FLAG antibody (HUABIO, China) at 4°C overnight with anti-GAPDH as loading control. Protein bands for *SpCas9* and GAPDH were captured by Bio-Rad ChemiDoc Touch. To determine stable *SpCas9* activity in primary liver tumor cell lines, we transfected these cells with the positive sgRNA control and performed targeted PCR amplicon and deep sequencing to evaluate editing efficiency on the targeted sites. Experiments were repeated three times in cell line 1C3-1 and done once in other cell lines.

### **Metaphase spread analysis of ploidy and chromosomal aberrations**

Cells were plated in 10 cm dish, grown to 70% confluence and treated with Colchicine in culture at 1.0 µg/mL for primary liver cancer cell line, 0.8 µg/mL for NIH-3T3 and Hepa1-6 cells and cells for 6 h. Cells were then harvested with 0.25% trypsin, resuspended in 75 mM KCl solution at 37 °C for 20-25 min and centrifuged at 1500 rpm for 5 min. After being fixed with methanol-acetic acid (3:1) for 15 min twice, the cells were dripped onto cold glass slides from 30 cm height and stained with 5% Giemsa for 10 min. Chromosome number and chromosomal aberration were assessed under light microscopy (400x). Images were analyzed by Image J.

### **Purification of genomic DNA and eccDNA**

gDNA was extracted from cells and 20 mg fresh liver tumor tissues using the Genomic DNA Miniprep Kit (AxyGEN, USA) according to the manufacturer's instructions. To isolate eccDNA from gDNA, Plasmid Safety ATP Dependent DNase (Lucigen, USA) was added into gDNA to remove linear DNA according to the manufacturer's manual as done previously<sup>28</sup>. We slightly modified the protocol to improve linear DNA digestion by adding 3 U DNase for 1 µg gDNA in a volume of 100 µL.

### **PCR amplicon deep sequencing of sgRNA target regions**

As done previously<sup>52</sup>, 35 targeted sites of gDNA were amplified by PCR with primers listed (**Table S7**). PCR products were purified by AxyPrep DNA Gel Extraction Kit or AxyPerp PCR Cleanup Kit (AxyGENE). Purified PCR products of each target were combined into a tube with equal volume. 1µg of PCR product mixture were used to establish DNA fragment libraries by VAHTS Universal DNA Library Prep Kit for Illumina V2 (Vazyme Biotech, China) and sequenced on Illumina HiSeq 2500.

## **Illumina sequencing data analysis, variant calling and clustering of variant frequencies**

Sequencing reads were aligned and processed according to Illumina sequencing analysis procedures and further analyzed by our analysis software developed previously<sup>33,52</sup>. Variant metrics including sequences and read depths was generated using in-house python scripts. Mutation frequency for each target gene in tissues and cells was calculated as total edited read count of the gene divided by total read count of the gene. To calculate allelic mutation ratio, the depth of sequence covering each mutation was divided by the overall read count of the corresponding gene. For each sample, we normalized read count of a specific gene by the library size to generate read ratios for gDNA and eccDNA separately. Data visualization was performed in R (v3.6.2) by ggplot2 package.

## **Sanger sequencing of eccDNA circularization junctions**

To identify circularization junction of eccDNA, outward PCR primers were designed and positioned on one side to TSG target site (**Table S7**). Outward PCR were performed using purified eccDNAs as a template. PCR products were verified first by DNA gel analysis. Visible PCR bands were purified and cloned into pUC19 vector and sequenced by Sanger sequencing. The sequences were analyzed to identify circularization junction of eccDNAs purified from 1C3-1 cells.

## **Declarations**

### **Ethics approval and consent to participate**

All animal procedures were reviewed and approved by the Animal Welfare and Ethical Review Body at Zhejiang University.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

All large number of microscope raw datasets and deep sequencing raw datasets from this work are available from the corresponding author on reasonable request, citing the experiments of interest.

## Competing interests

The authors declare no conflicts of interests

## Funding

This work is funded by the National Natural Science Foundation of China (No. 31870806 and No. 31671385 to A.Y.X., and No. 32071439 to Y.L.F), the Department of Science and Technology of Zhejiang Province (2015C03047 to A.Y.X.), and the Natural Science Foundation of Zhejiang Province (LQ20C050004 to S.C.L).

## Authors' contributions

AYX conceived of the study and designed the experiments. AYX and HL supervised the study. TG, GQC, MW, XYY and KML conducted the experiments. TG, XFL, PYL and AYX performed bioinformatic analysis. SCL and YLF assisted with experiments and bioinformatic analysis, and YLF performed statistical analysis. AYX, TG, GQC, XFL and PYL analyzed the results. TG, AYX and XFL prepared figures and tables, and AYX and TG wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

The authors are grateful to the other members of the Xie laboratory for helpful discussions.

## References

1. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**, 585–598 (2014).
2. Davis, A., Gao, R. & Navin, N. Tumor evolution: Linear, branching, neutral or punctuated? *Biochim Biophys Acta Rev Cancer* **1867**, 151–161 (2017).
3. Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* **37**, 471–484 (2020).
4. Vendramin, R., Litchfield, K. & Swanton, C. Cancer evolution: Darwin and beyond. *EMBO J* **40**, e108389 (2021).
5. Baca, S. C. *et al.* Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
6. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **52**, 331–341 (2020).

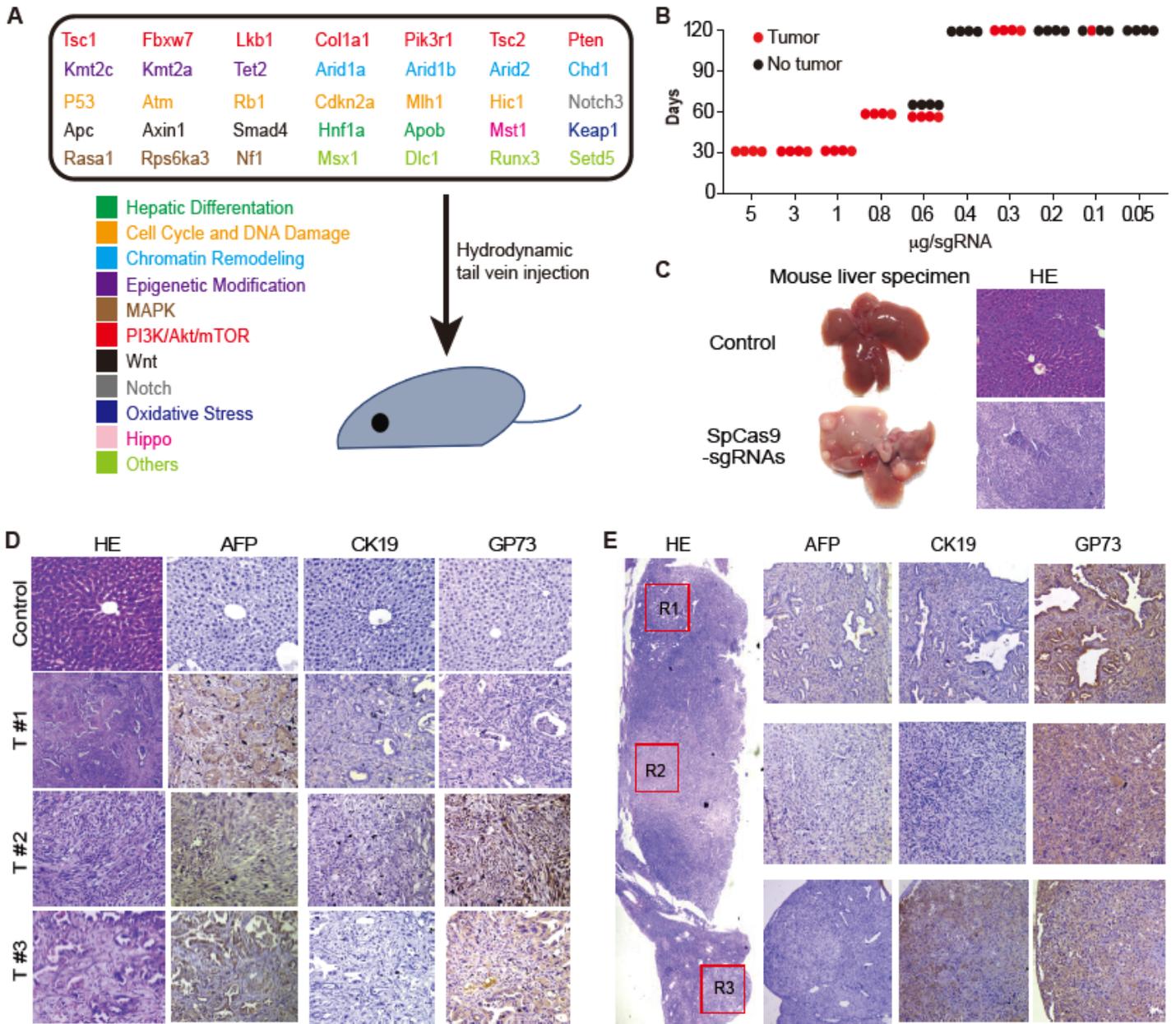
7. Ly, P. & Cleveland, D. W. Rebuilding Chromosomes After Catastrophe: Emerging Mechanisms of Chromothripsis. *Trends Cell Biol* **27**, 917–930 (2017).
8. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
9. Kloosterman, W. P. *et al.* Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biol* **12**, R103 (2011).
10. Ling, S. *et al.* Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proc Natl Acad Sci U S A* **112**, E6496-6505 (2015).
11. Malhotra, A. *et al.* Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res* **23**, 762–776 (2013).
12. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589–593 (2012).
13. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
14. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat Genet* **47**, 209–216 (2015).
15. Craig, A. J., von Felden, J., Garcia-Lezana, T., Sarcognato, S. & Villanueva, A. Tumour evolution in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol* **17**, 139–152 (2020).
16. Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **7**, 6 (2021).
17. Brown, Z. J., Heinrich, B. & Greten, T. F. Mouse models of hepatocellular carcinoma: an overview and highlights for immunotherapy research. *Nat Rev Gastroenterol Hepatol* **15**, 536–554 (2018).
18. Weber, J. & Rad, R. Engineering CRISPR mouse models of cancer. *Curr Opin Genet Dev* **54**, 88–96 (2019).
19. Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380–384 (2014).
20. Song, C.-Q. *et al.* Genome-Wide CRISPR Screen Identifies Regulators of Mitogen-Activated Protein Kinase as Suppressors of Liver Tumors in Mice. *Gastroenterology* **152**, 1161-1173.e1 (2017).
21. Wang, G. *et al.* Mapping a functional cancer genome atlas of tumor suppressors in mouse liver using AAV-CRISPR-mediated direct in vivo screening. *Sci Adv* **4**, eaao5508 (2018).
22. Weber, J. *et al.* CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13982–13987 (2015).
23. Zhu, M. *et al.* Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell* **177**, 608-621.e12 (2019).
24. Jasin, M. & Haber, J. E. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst.)* **44**, 6–16 (2016).
25. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat Genet* **53**, 895–905 (2021).

26. Møller, H. D. *et al.* CRISPR-C: circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Res* **46**, e131 (2018).
27. Bailey, C., Shoura, M. J., Mischel, P. S. & Swanton, C. Extrachromosomal DNA-relieving heredity constraints, accelerating tumour evolution. *Ann Oncol* **31**, 884–893 (2020).
28. Turner, K. M. *et al.* Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**, 122–125 (2017).
29. Verhaak, R. G. W., Bafna, V. & Mischel, P. S. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer* **19**, 283–288 (2019).
30. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet* **52**, 891–897 (2020).
31. Wu, S. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
32. Zhu, Y. *et al.* Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* **39**, 694-707.e7 (2021).
33. Feng, Y.-L. *et al.* H2AX facilitates classical non-homologous end joining at the expense of limited nucleotide loss at repair junctions. *Nucleic Acids Res.* **45**, 10614–10633 (2017).
34. Liu, F., Song, Y. & Liu, D. Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Ther* **6**, 1258–1266 (1999).
35. Loeb, L. A. A mutator phenotype in cancer. *Cancer Res* **61**, 3230–3239 (2001).
36. Hintzsche, H. *et al.* Fate of micronuclei and micronucleated cells. *Mutat Res Rev Mutat Res* **771**, 85–98 (2017).
37. Darlington, G. J., Bernhard, H. P., Miller, R. A. & Ruddle, F. H. Expression of liver phenotypes in cultured mouse hepatoma cells. *J Natl Cancer Inst* **64**, 809–819 (1980).
38. Paulsen, T., Kumar, P., Koseoglu, M. M. & Dutta, A. Discoveries of Extrachromosomal Circles of DNA in Normal and Tumor Cells. *Trends Genet.* **34**, 270–278 (2018).
39. Donne, R., Saroul-Ainama, M., Cordier, P., Celton-Morizur, S. & Desdouets, C. Polyploidy in liver development, homeostasis and disease. *Nat Rev Gastroenterol Hepatol* **17**, 391–405 (2020).
40. Cox, D., Yuncken, C. & Spriggs, A. I. MINUTE CHROMATIN BODIES IN MALIGNANT TUMOURS OF CHILDHOOD. *Lancet* **1**, 55–58 (1965).
41. Cai, M. *et al.* Inhibiting homologous recombination decreases extrachromosomal amplification but has no effect on intrachromosomal amplification in methotrexate-resistant colon cancer cells. *Int J Cancer* **144**, 1037–1048 (2019).
42. Carroll, S. M. *et al.* Characterization of an episome produced in hamster cells that amplify a transfected CAD gene at high frequency: functional evidence for a mammalian replication origin. *Mol Cell Biol* **7**, 1740–1750 (1987).
43. Ly, P. *et al.* Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat Genet* **51**, 705–715 (2019).

44. Tanaka, H. & Yao, M.-C. Palindromic gene amplification—an evolutionarily conserved role for DNA inverted repeats in the genome. *Nat. Rev. Cancer* **9**, 216–224 (2009).
45. Wang, Y. *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* **599**, 308–314 (2021).
46. Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
47. Umbreit, N. T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science* **368**, eaba0712 (2020).
48. Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Mol Syst Biol* **11**, 828 (2015).
49. Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D. & Regenberg, B. Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci U S A* **112**, E3114-3122 (2015).
50. Shibata, Y. *et al.* Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* **336**, 82–86 (2012).
51. Nathanson, D. A. *et al.* Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science* **343**, 72–76 (2014).
52. Guo, T. *et al.* Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol.* **19**, 170 (2018).

## Figures

**Figure 1**



**Figure 1**

**Induction of primary mouse liver cancer by CRISPR/Cas9-mediated somatic mutagenesis in mice.**

(A) 34 TSGs and the Setd5 control site targeted by multiplexed CRISPR/Cas9. Each TSG is indicated with its related cancer signaling pathway. Hepatic delivery of *SpCas9* and sgRNA expression plasmids is achieved by hydrodynamic tail vein injection (HTVI).

(B) Effect of sgRNA dosage on CRISPR/Cas9-induced liver tumor formation. Each dot indicates one mouse.

(C) Mouse liver specimen with or without tumor nodules (left) and H&E staining of a tumor nodule (right).

(D) Microscopic IHC images of CRISPR/Cas9-induced liver tumors. T #1, T #2 and T #3 represent tumor nodules from three mice. AFP and GP73: HCC biomarkers; CK19: ICC biomarker.

(E) Histology and IHC staining of a liver tumor nodule indicating three heterogeneous regions. Left: these three regions indicated with R1, R2 and R3 in H&E staining; Right: microscopic IHC images for these three regions.

Figure 2

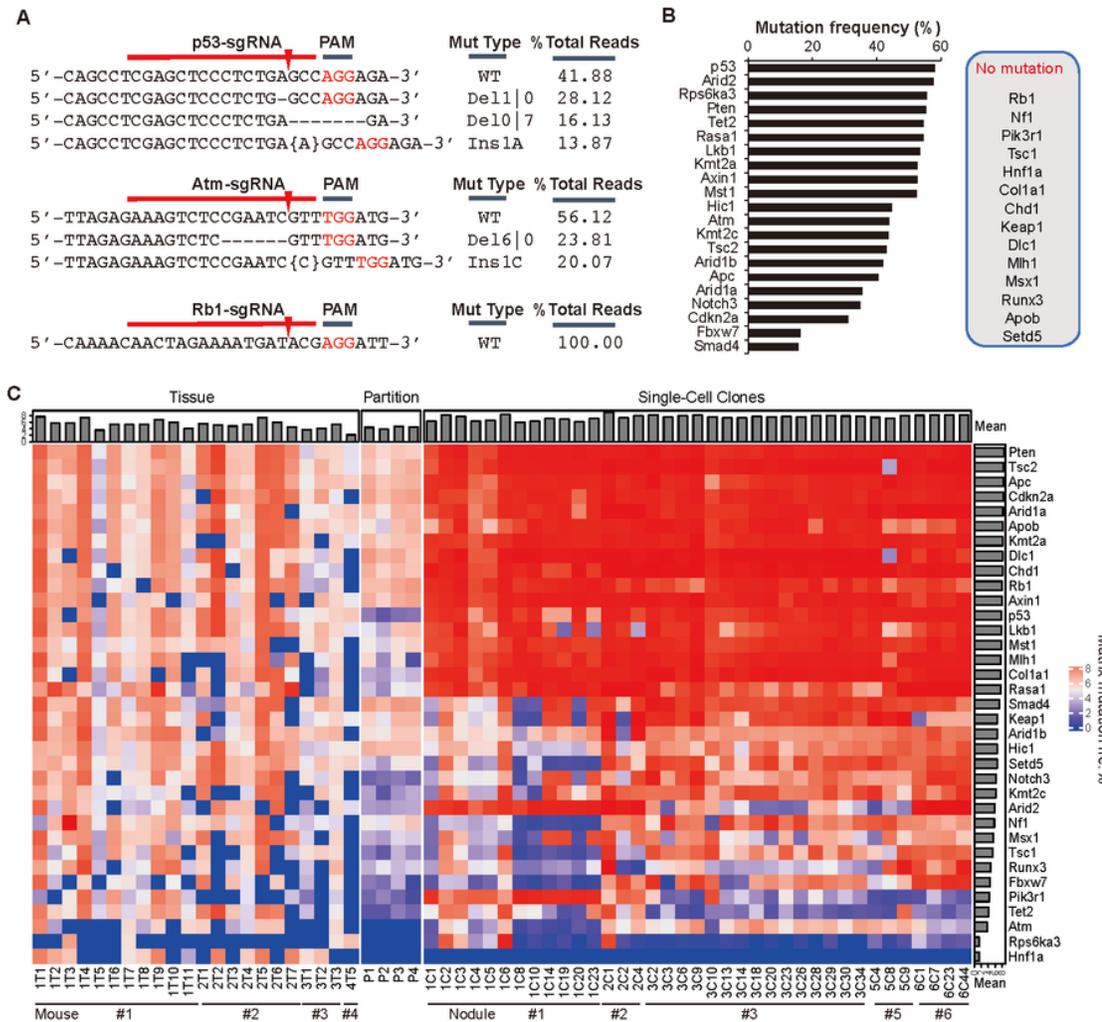


Figure 2

Target TSG mutation profiles in CRISPR/Cas9-induced mouse liver tumor nodules and single-cell clones

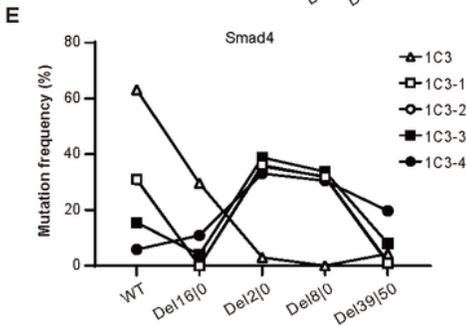
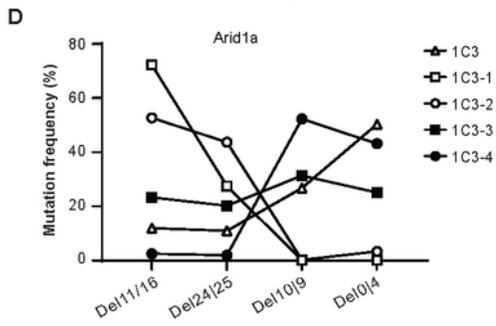
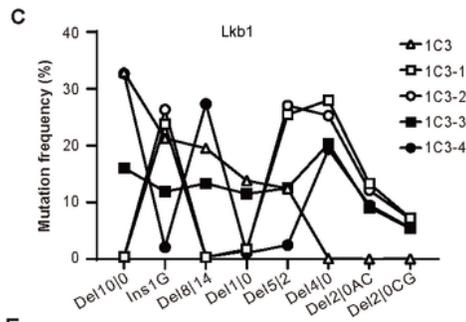
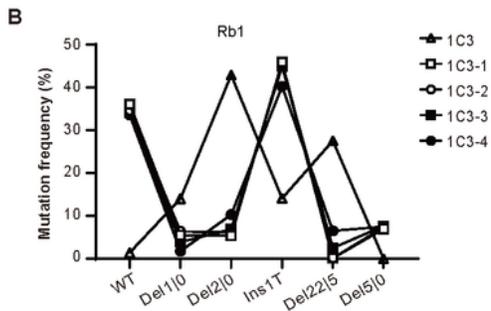
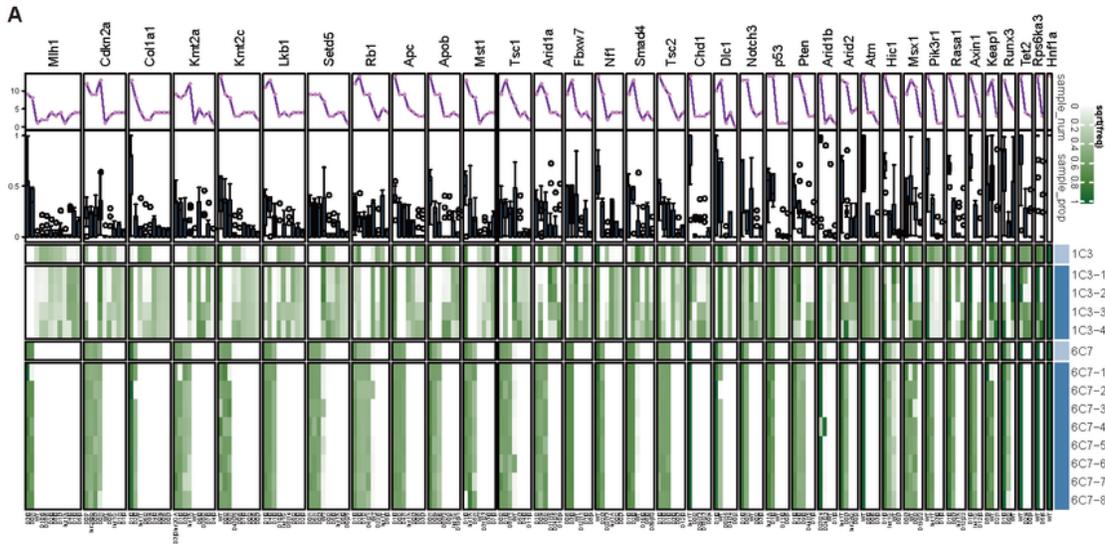
(A) Representative target mutation types of *p53*, *Atm* and *Rb1* in a tumor nodule and the frequencies of these mutations. sgRNA target sites with PAM and the cleavage position are indicated. Dash lines and nucleotides in parenthesis in the sequences denote deletion and inserted nucleotides, respectively. WT,

wild-type; Del, deletion; Ins, insertion. The numbers flanking Del or Ins indicate the numbers of deleted or inserted nucleotides.

**(B)** Representative target mutations of 21 TSGs with cumulative frequencies in a tumor nodule. 14 unedited gene are also listed on the right.

**(C)** Mutational landscape of 35 targets in 22 primary tumor nodules, 4 parts from a single tumor nodule and 37 single-cell clones as indicated. Top and right bars indicate the mean of the mutation frequencies for each sample and each gene, respectively. Cumulative matrix mutation frequencies are highlighted from the highest 8 in red to the lowest 0 in blue.

**Figure 3**



**Figure 3**

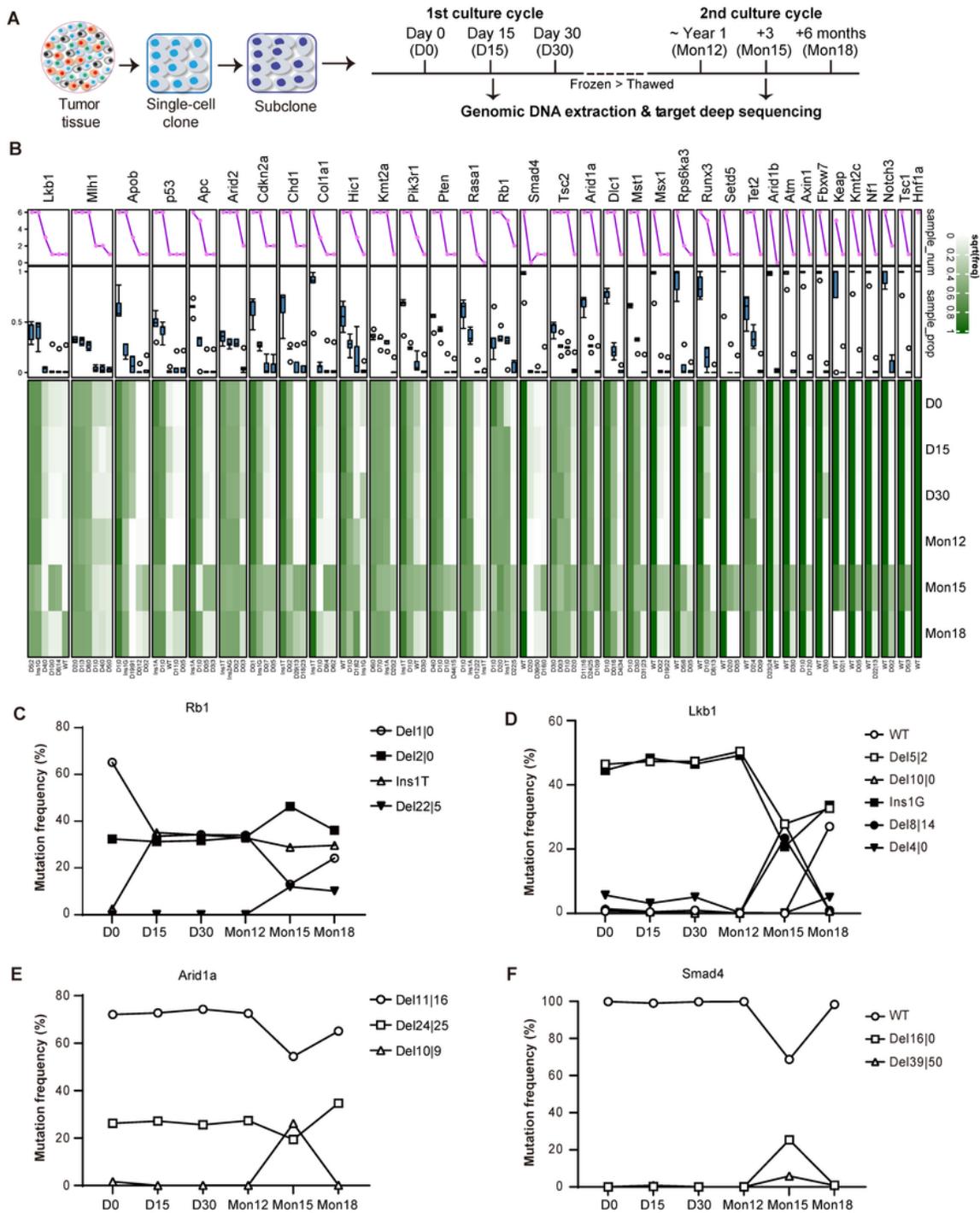
**Alteration of target site mutations between subclones of single-cell clone 1C3 and 6C7 derived from mouse liver tumors**

(A) Changes in target site mutation profiles between parental single-cell clones (i.e., 1C3 and 6C7) and their subclones (1C3-1, -2, -3 and -4 for 1C3 and 6C7-1, -2, -3, -4, -5, -6, -7 and -8 for 6C7). Gene targets and

their mutation types are shown on top and at bottom, respectively. The numbers flanking Del or Ins indicated at bottom are the numbers of deleted or inserted nucleotides. Sqrt(freq) representing square root of mutation frequencies is indicated from the highest 1 in dark green to the lowest 0 in white. The top line chart and the boxplot under show the number of single-cell clones with a specific mutation type (i.e., sample\_num) and the percentage distribution of a specific mutation type in all single-cell clones (i.e., sample\_prop), respectively.

**(B-E)** Representative target mutation oscillation of *Rb1* **(B)**, *Lkb1* **(C)**, *Arid1a* **(D)** and *Smad4* **(E)** between parental clone 1C3 and its subclones.

**Figure 4**



**Figure 4**

Oscillation of target site mutations during proliferation of the single-cell subclone 1C3-1 derived from 1C3.

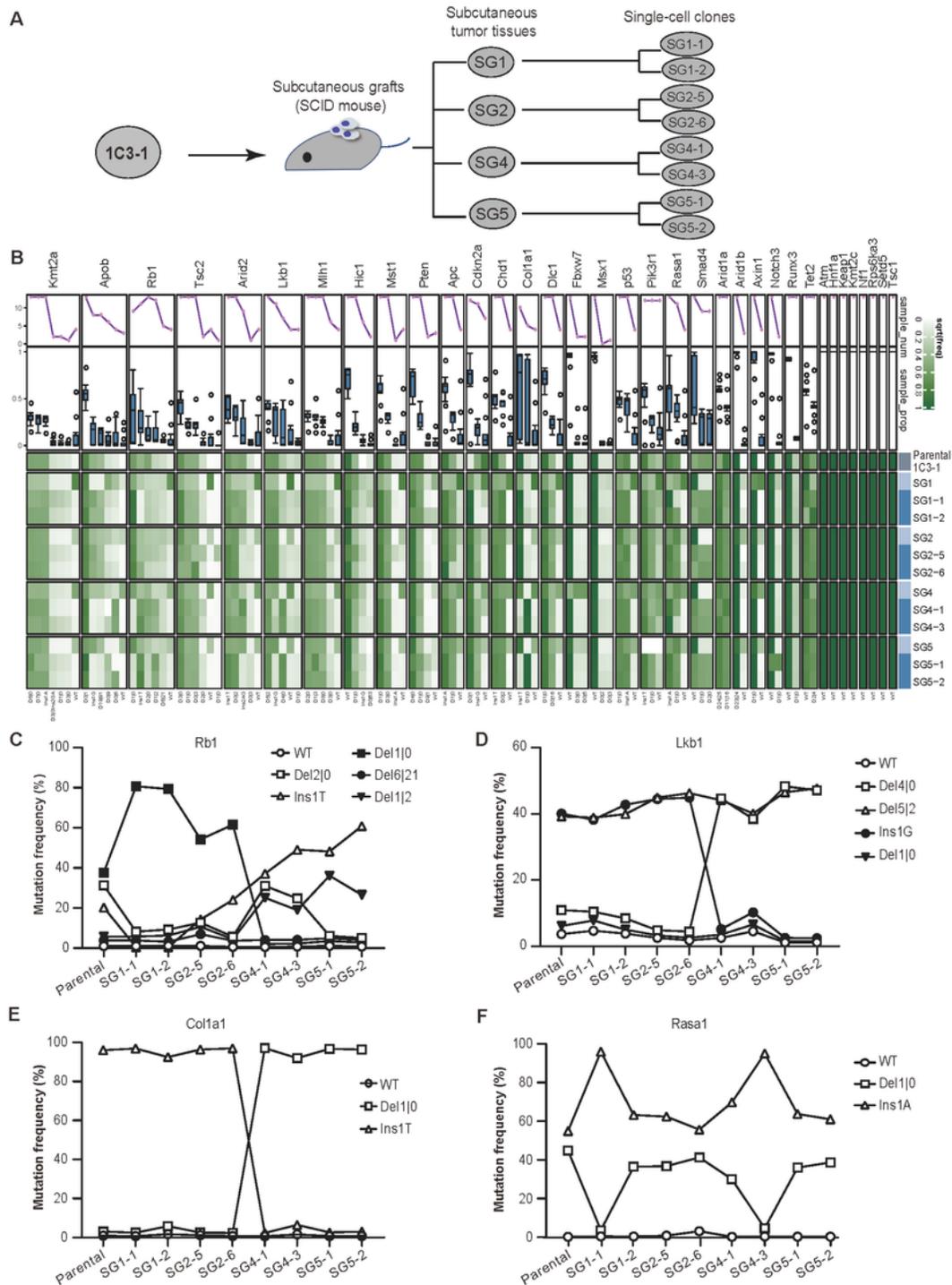
(A) Schematic of experimental outline. Proliferation of the single-cell clone 1C3-1 was divided the first culture cycle that continues for 30 days and the second culture cycle that continues after 1 year storage.

Target site mutations were determined by targeted PCR amplicon deep sequencing of gDNA extracted from cells at different time points of cell proliferation as indicated. D0, Day 0; D15, Day 15; D30, Day 30; Mon12, Year 1; Mon15, Year 1+3 months; Mon12, Year 1+6 months.

**(B)** Changes in target site mutation profiles of 1C3-1 at indicated time points of cell proliferation (i.e., Day 0, 15, 30, and Mon 12, 15, and 18). Gene targets and their mutation types are shown on top and at bottom, respectively. The numbers flanking Del or Ins indicated at bottom are the numbers of deleted or inserted nucleotides. Sqrt(freq) representing square root of mutation frequencies is indicated from the highest 1 in dark green to the lowest 0 in white. The top line chart and the boxplot under show the number of single-cell clones with a specific mutation type (i.e., sample\_num) and the percentage distribution of a specific mutation type in all single-cell clones (i.e., sample\_prop), respectively.

**(C-F)** Representative target mutation oscillation of *Rb1* **(B)**, *Lkb1* **(C)**, *Arid1a* **(D)** and *Smad4* **(E)** during cell proliferation of 1C3-1.

**Figure 5**



**Figure 5**

**Oscillation of target site mutations during subcutaneous tumor cell grafts of 1C3-1.**

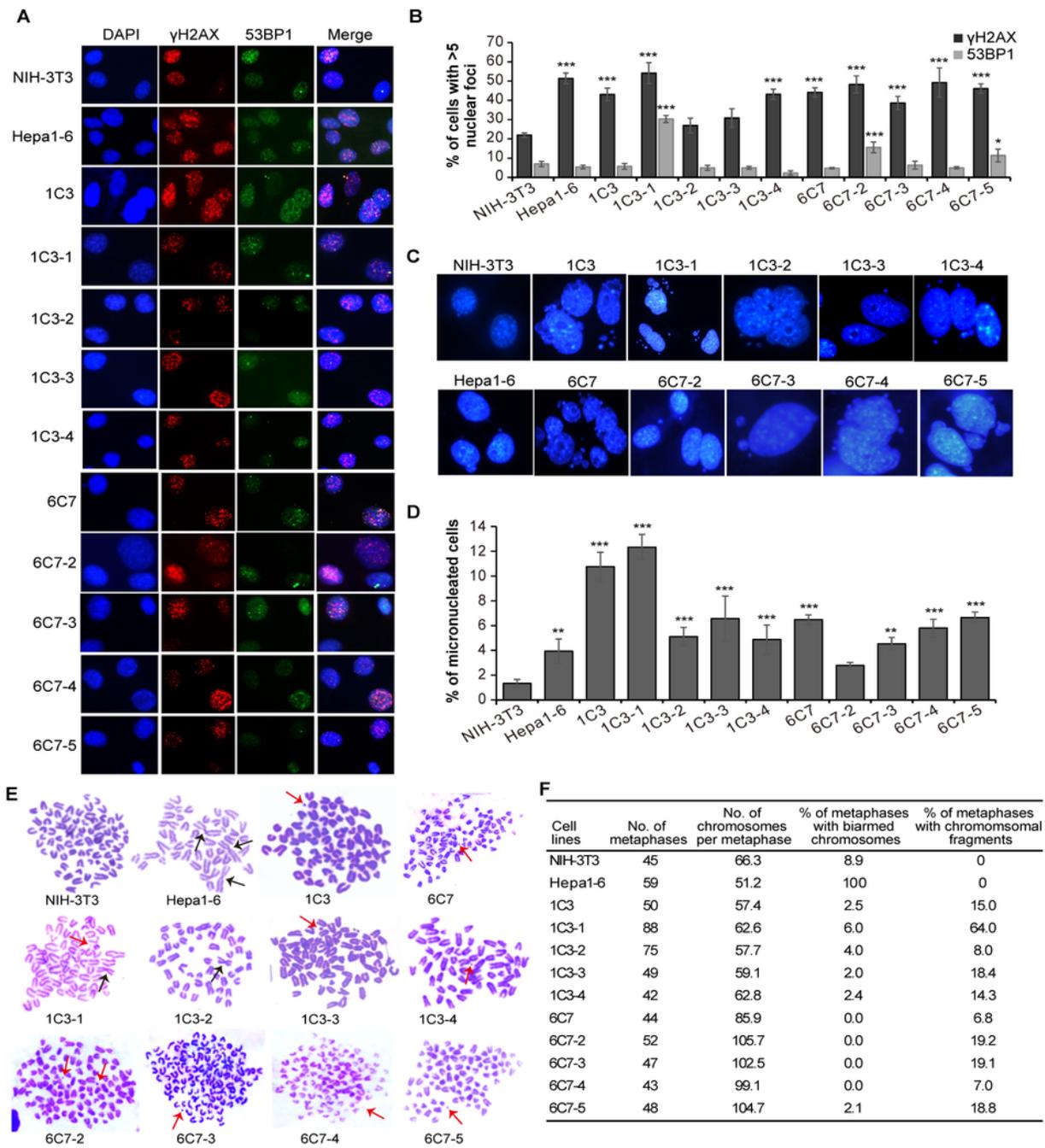
(A) Schematic of experimental outline. After subcutaneous implantation of the single-cell subclone 1C3-1 into SCID mice, tumors were formed and four tumor tissues (SG1, SG2, SG4 and SG5) from different mice collected at 2 weeks post inoculation. Two single-cell clones were derived from each of these four tumor

tissues as indicated. Target site mutations were determined by targeted PCR amplicon deep sequencing of gDNA extracted from cells. SG, subcutaneous graft.

**(B)** Changes in target site mutation profiles of single-cell clones from subcutaneous tumor cell grafts as indicated. Gene targets and their mutation types are shown on top and at bottom, respectively. The numbers flanking Del or Ins indicated at bottom are the numbers of deleted or inserted nucleotides. Sqrt(freq) representing square root of mutation frequencies is indicated from the highest 1 in dark green to the lowest 0 in white. The top line chart and the boxplot under show the number of single-cell clones with a specific mutation type (i.e., *sample\_num*) and the percentage distribution of a specific mutation type in all single-cell clones (i.e., *sample\_prop*), respectively.

**(C-F)** Representative target mutation oscillation of *Rb1* **(B)**, *Lkb1* **(C)**, *Col1a1* **(D)** and *Rasa1* **(E)** across single-cell clones derived from subcutaneous tumor cell grafts.

**Figure 6**



**Figure 6**

Genomic instability in single-cell clones derived from CRISPR/Cas9-induced mouse liver tumors.

(A) Microscopic images of spontaneous γH2AX and 53BP1 foci in primary mouse liver cancer cells.

**(B)** Percentage of cells with >5  $\gamma$ H2AX foci or 53BP1 foci. Columns indicate the mean  $\pm$  S.E.M from three independent experiments with statistical significance detected by One-way ANOVA with Dunnett's multiple comparison test (vs. NIH-3T3). \*,  $P < 0.05$  and \*\*\*,  $P < 0.001$ .

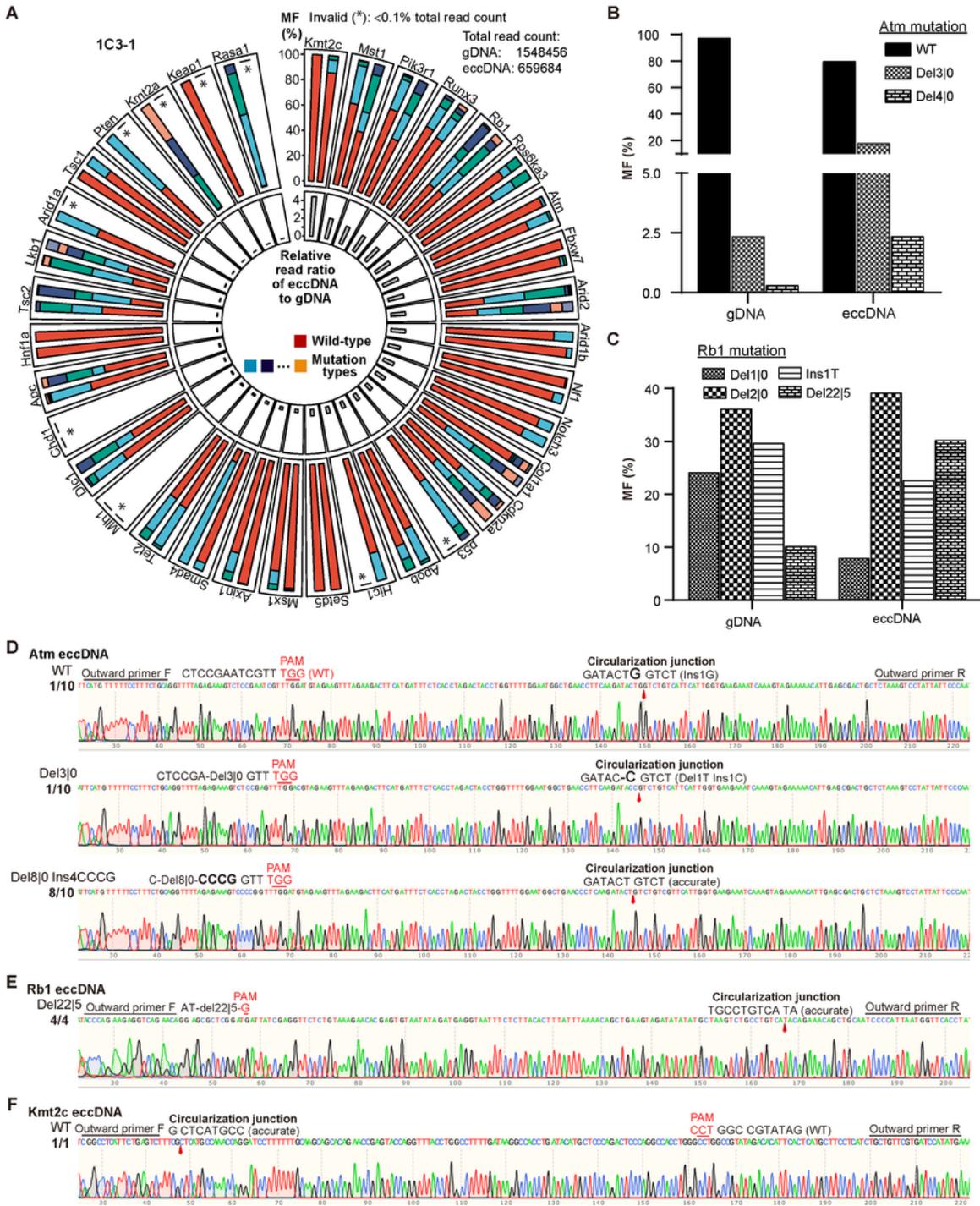
**(C)** Microscopic images of micronuclei (MN) in primary mouse liver cancer cells.

**(D)** Percentage of cells with micronuclei. Columns indicate the mean  $\pm$  S.E.M from three independent experiments with statistical significance detected by One-way ANOVA with Dunnett's multiple comparison test (vs. NIH-3T3). \*\*,  $P < 0.01$  and \*\*\*,  $P < 0.001$ .

**(E)** Representative images of chromosomal aberrations in metaphase spread of primary mouse liver cancer cell lines. Red arrows and black arrows indicate chromosomal fragments and biarmed chromosomes, respectively.

**(F)** Summary of metaphase spread analysis. In each indicated cell line, the number of metaphases, number of chromosomes per metaphase, percentages of metaphases with biarmed chromosomes and with chromosomal fragments are calculated and shown.

**Figure 7**



**Figure 7**

**Target site mutations of TSGs detected on eccDNA**

(A) Circular barplot for target site mutations of 35 targets in gDNA and eccDNA in representative single-cell clone 1C3-1. The column in the inner circle for each target gene were relative read ratios of eccDNA to gDNA for each target site from deep sequencing data. The stacked barplot in the outer circle represents

the mutation pattern at indicated target sites in eccDNA and gDNA. Mutation types are shown in color. Total read counts of eccDNA and gDNA are listed. \*, invalid read counts at indicated sites, which are defined as <0.1% of total read counts from eccDNA or gDNA.

**(B,C)** Target site mutation distribution of representative TSGs *Atm* **(B)** and *Rb1* **(C)** in gDNA and eccDNA.

**(D-F)** Circularization junctions of eccDNA containing target sites for *Atm* **(D)**, *Rb1* **(E)** and *Kmt2c* **(F)**. Circularization junction sequences as shown are determined by PCR with a pair of indicated outward primers followed by Sanger sequencing. Target sites with PAM and mutations are also indicated.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.ai](#)
- [FigureS2.ai](#)
- [FigureS3.ai](#)
- [FigureS4.ai](#)
- [FigureS5.ai](#)
- [FigureS6.ai](#)
- [FigureS7.ai](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable7.xlsx](#)
- [Supplementaryinformation.docx](#)