7-9-2020

Genes adapt to outsmart gene-targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation.

Vishnu Hosur

Benjamin E. Low

Daniel Li

Grace Stafford

Vivek Kohar

See next page for additional authors

Authors

Vishnu Hosur, Benjamin E. Low, Daniel Li, Grace Stafford, Vivek Kohar, Leonard D. Shultz, and Michael V. Wiles

Check for
updates

Genes adapt to outsmart gene-targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation

Vishnu Hosur[1][*] , Benjamin E. Low[1], Daniel Li[2], Grace A. Stafford[1], Vivek Kohar[1], Leonard D. Shultz[1] and Michael V. Wiles[1][*]

* Correspondence: vishnu.hosur@
jax.org; michael.wiles@jax.org
[1]The Jackson Laboratory for
Mammalian Genetics, Bar Harbor,
ME, USA
Full list of author information is
available at the end of the article

Abstract

Background: Gene disruption in mouse embryonic stem cells or zygotes is a conventional genetics approach to identify gene function in vivo. However, because different gene disruption strategies use different mechanisms to disrupt genes, the strategies can result in diverse phenotypes in the resulting mouse model. To determine whether different gene disruption strategies affect the phenotype of resulting mutant mice, we characterized *Rhbdf1* mouse mutant strains generated by three commonly used strategies—definitive-null, targeted knockout (KO)-first, and CRISPR/Cas9.

Results: We find that *Rhbdf1* responds differently to distinct KO strategies, for example, by skipping exons and reinitiating translation to potentially yield gain-of-function alleles rather than the expected null or severe hypomorphic alleles. Our analysis also revealed that at least 4% of mice generated using the KO-first strategy show conflicting phenotypes.

Conclusions: Exon skipping is a widespread phenomenon occurring across the genome. These findings have significant implications for the application of genome editing in both basic research and clinical practice.

Keywords: Definitive-null, Knockout-first, CRISPR/Cas9, Gene disruption, Exon skipping, Gene editing, And RHBDF1

Background

Mice are closely related, genetically, to humans, and hence mice have been chosen as a model system to decipher the function of ~ 20,000 protein-coding genes to gain insights into human biology and disease. For large-scale mouse mutagenesis efforts, gene targeting by homologous recombination in mouse embryonic stem (ES) cells was an efficient and versatile technique. Gene targeting involved either a definitive-null design, which deleted the entire genomic sequence of the target gene, or a targeted knockout

Hosur *et al. Genome Biology*　　(2020) 21:168

Page 2 of 25

(KO)-first design, which offered several advantages, including a gene disruption and a reporter-tagged mutation, and furthermore, permitted analysis of gene function in a tissue-specific or temporal manner. Recently, the use of CRISPR/Cas9 to disrupt genes directly in zygotes has superseded both the definitive-null and KO-first strategies. To determine whether different gene-targeting strategies affect the phenotype of homozygous mutant mice, we systematically characterize *Rhbdf1* mutant mice generated by these three KO strategies—definitive-null, targeted KO-first, and CRISPR/Cas9. The *Rhbdf1* gene encodes RHBDF1 and has been implicated as playing a key role in growth and development [1], inflammation [2], and cancer [3–5].

Definitive-null and targeted KO-first strategies are powerful high-throughput approaches that can be used for large-scale gene targeting in ES cells to study thousands of mammalian protein-coding genes to better understand human biology and disease [6–8]. In the use of the definitive-null strategy, bacterial artificial chromosome (BAC)-based targeting vectors replace the entire genomic sequence of the target gene (Additional file 1: Figure S1a), thereby generating a null allele. In contrast, the targeted KO-first approach [9, 10], a strategy that includes alternative steps that can be chosen based on the desired result, is highly versatile, as it enables the analysis of gene function in a tissue-specific or temporal manner. The KO-first allele is created through targeted delivery of a cassette that includes a mouse *En2* splice site, a β-galactosidase reporter, a neomycin resistance gene, and a transcriptional termination SV40 polyadenylation sequence, into the intronic sequence. Transcription of the targeted gene is terminated by the polyadenylation sequence before the entire gene sequence is transcribed, resulting in truncated transcription of a nonfunctional reporter protein. Application of Cre recombinase to a KO-first ES cell clone results in generation of a null allele (Additional file 1: Figure S1b). Application of Flp recombinase to a KO-first ES cell clone converts the KO-first allele into a conditional-ready (or simply a wildtype) allele by removing the cassette flanked by the *FRT* sites (Additional file 1: Figure S1c); subsequent application of Cre recombinase elegantly generates conditional alleles by deleting exon(s) that are flanked by *loxP* sites (Additional file 1: Figure S1c). The "KO-first" alleles, by offering substantial flexibility with respect to either gene disruption or target tissue expression and timing, are time-efficient. Despite evidence suggesting that targeted KO-first alleles may result in hypomorphs [11], to our knowledge there have been no reports indicating that use of the targeted KO-first strategy has led to gain-of-function alleles that rescue the severe phenotype caused by the null alleles.

CRISPR/Cas9-mediated genome editing is currently the primary conventional approach for generating null alleles [12]. This method relies on the introduction of frameshift mutations via non-homologous end joining (NHEJ) in ES cells or zygotes to trigger nonsense-mediated decay (NMD) of mRNA transcripts [13]. However, emerging evidence suggests that these mutant alleles can reinitiate translation from a downstream start codon resulting in truncated proteins [14]. Recently, two separate studies performed systematic analysis of CRISPR/Cas9-induced frameshift mutations in HAP1 cells and revealed that one third of the genes could generate truncated proteins either by reinitiating translation or skipping the edited exon [15, 16]. Likewise, in zebrafish, CRISPR/Cas9-mediated generation of KO alleles resulted in unpredicted mRNA transcripts owing to exon skipping, nonsense-associated altered splicing, or use of cryptic

Hosur *et al. Genome Biology*   (2020) 21:168

Page 3 of 25

splice sites [17]. Collectively, these studies highlight the limitations of CRISPR/Cas9-mediated generation of null alleles.
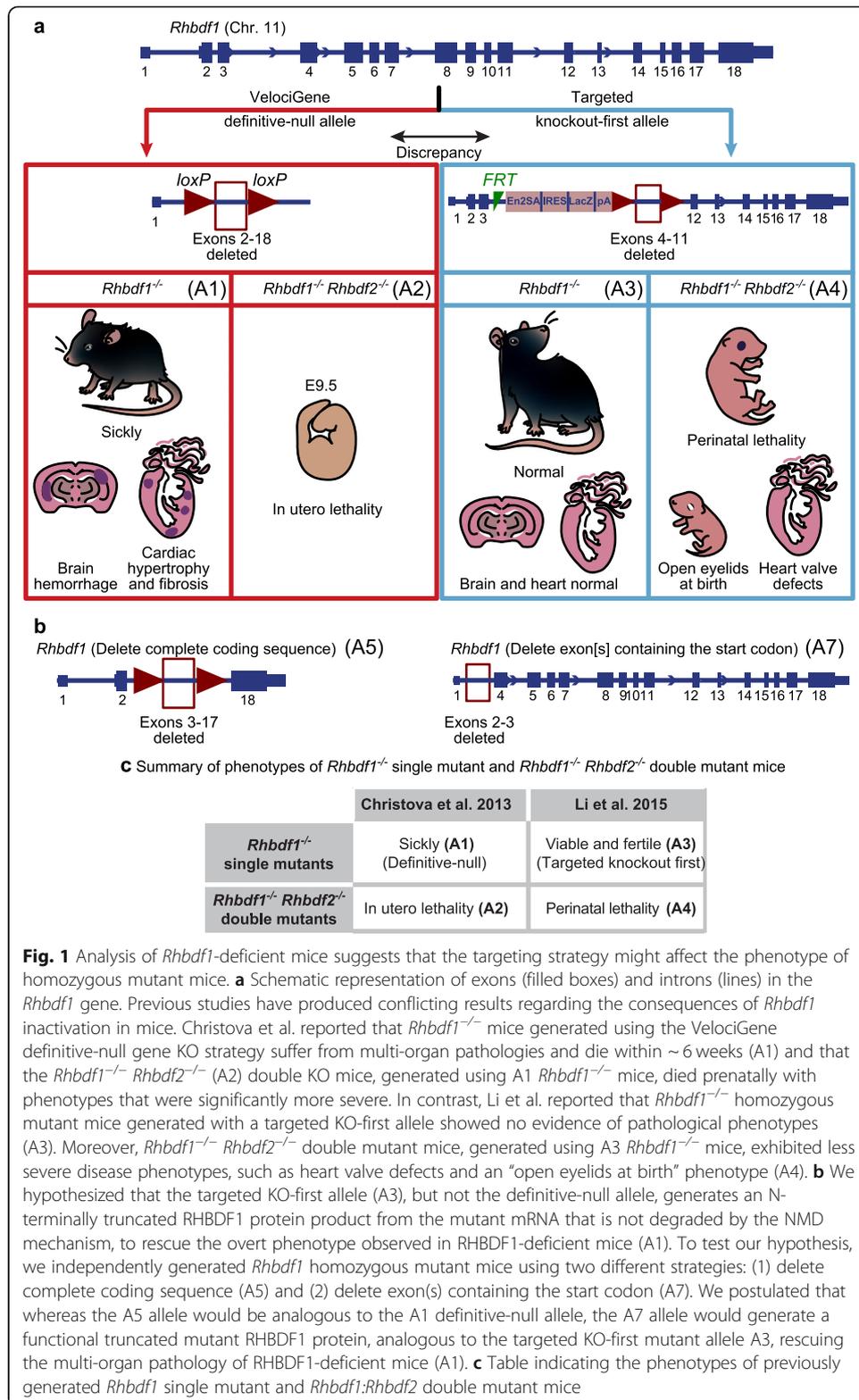
Building on the aforementioned work in HAP1 cell lines and mutant zebrafish lines, we comprehensively characterized *Rhbdf1* mutant mouse strains generated by three distinct gene-targeting strategies—definitive-null, targeted KO-first, and CRISPR/Cas9—using whole-exome sequencing (Exome-Seq), RNA sequencing (RNA-Seq), 5′-rapid amplification of cDNA ends (RACE), and functional analysis of residual proteins. We find that whereas *definitive-null Rhbdf1* mutant mice (hereafter referred to as *Rhbdf1*$^{null/null}$) show growth retardation, brain and heart defects, and the death of the majority of mice by postnatal day 14 (P14), targeted KO-first- and CRISPR/Cas9-engineered *Rhbdf1* mutant mice do not display growth retardation or any of the other pathologies observed in the *Rhbdf1*$^{null/null}$ mice, suggesting incomplete elimination of RHBDF1 biological activity. Comprehensive analyses of all three *Rhbdf1* mutant alleles revealed the existence of functional N-terminally truncated RHBDF1 mutant proteins capable of rescuing the multi-organ pathology observed in *Rhbdf1*$^{null/null}$ mice. Our data suggest that because mutant transcripts are capable of generating functional truncated proteins, a thorough post-genomic-DNA analysis of targeted KO-first- and CRISPR/Cas9-engineered alleles, and additional studies as well, is warranted to validate the null alleles before phenotypic characterization of mutant mouse strains.

Results

Does the choice of gene-targeting strategy affect the phenotype?

In recent studies, two different groups each generated *Rhbdf1* homozygous mutant mice, and reported conflicting results regarding the phenotypes [1, 18]; Christova and colleagues, using the definitive-null strategy, generated an *Rhbdf1* loss-of-function mutation in two different inbred mouse strains: C57BL/6J and 129S6. In both strains, heterozygous-null mice were normal and indistinguishable from wildtype littermates, whereas homozygous-null mice had low body weights and died at about 9 days (C57BL/6J) and 6 weeks (129S6) after birth (Fig. 1a, see A1). In contrast, using the targeted KO-first strategy, Li and colleagues reported that *Rhbdf1* mutant mice generated on a mixed genetic background—C57BL/6N and FVB—were healthy with no apparent phenotypic or histological abnormalities, and proposed that RHBDF1 is not essential for mouse development (Fig. 1a, see A3). Furthermore, Christova et al. (using both C57BL/6J and 129S6 mice) and Li et al. both generated *Rhbdf1:Rhbdf2* double mutant mice, and the phenotypes differed considerably between the two studies. While Christova et al. reported in utero lethality in the double mutants (Fig. 1a, see A2), Li et al. observed perinatal lethality with an "eyelids open at birth" phenotype and heart valve defects (Fig. 1a, see A4).

It seems unlikely that there is a modifier in the B6 congenic E2a-Cre strain [B6.FVB-TgEIIa-cre; Stock No: 003724] strain that could account for the difference in the phenotype observed in the single and double mutant mice generated on the C57BL/6J background by Christova et al. It seems more likely, rather, that the differences in the phenotypes could be due to effects of the different gene-targeting strategies employed (Fig. 1a) by the two studies to generate *Rhbdf1* KO mice. Notably, both the transcription start site and the start codon remain unaltered by the targeted KO-first strategy.
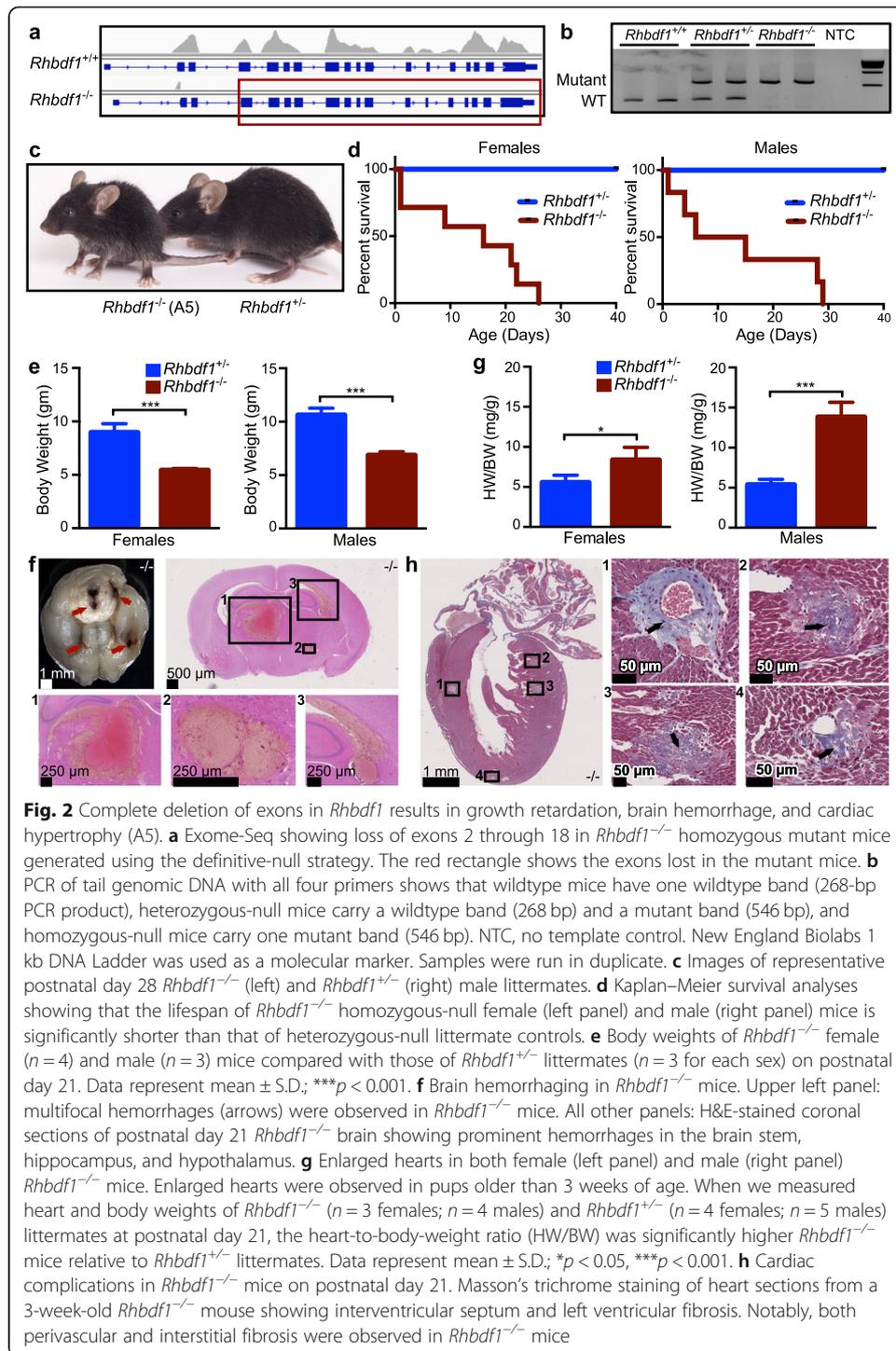
Fig. 1 Analysis of *Rhbdf1*-deficient mice suggests that the targeting strategy might affect the phenotype of homozygous mutant mice. **a** Schematic representation of exons (filled boxes) and introns (lines) in the *Rhbdf1* gene. Previous studies have produced conflicting results regarding the consequences of *Rhbdf1* inactivation in mice. Christova et al. reported that *Rhbdf1*$^{-/-}$ mice generated using the VelociGene definitive-null gene KO strategy suffer from multi-organ pathologies and die within ~ 6 weeks (A1) and that the *Rhbdf1*$^{-/-}$ *Rhbdf2*$^{-/-}$ (A2) double KO mice, generated using A1 *Rhbdf1*$^{-/-}$ mice, died prenatally with phenotypes that were significantly more severe. In contrast, Li et al. reported that *Rhbdf1*$^{-/-}$ homozygous mutant mice generated with a targeted KO-first allele showed no evidence of pathological phenotypes (A3). Moreover, *Rhbdf1*$^{-/-}$ *Rhbdf2*$^{-/-}$ double mutant mice, generated using A3 *Rhbdf1*$^{-/-}$ mice, exhibited less severe disease phenotypes, such as heart valve defects and an "open eyelids at birth" phenotype (A4). **b** We hypothesized that the targeted KO-first allele (A3), but not the definitive-null allele, generates an N-terminally truncated RHBDF1 protein product from the mutant mRNA that is not degraded by the NMD mechanism, to rescue the overt phenotype observed in RHBDF1-deficient mice (A1). To test our hypothesis, we independently generated *Rhbdf1* homozygous mutant mice using two different strategies: (1) delete complete coding sequence (A5) and (2) delete exon(s) containing the start codon (A7). We postulated that whereas the A5 allele would be analogous to the A1 definitive-null allele, the A7 allele would generate a functional truncated mutant RHBDF1 protein, analogous to the targeted KO-first mutant allele A3, rescuing the multi-organ pathology of RHBDF1-deficient mice (A1). **c** Table indicating the phenotypes of previously generated *Rhbdf1* single mutant and *Rhbdf1:Rhbdf2* double mutant mice

To address the underlying cause of discrepancy, we generated (a) *Rhbdf1* KO mice using *the definitive-null design*, which results in alleles that delete the entire protein-coding sequence of the target gene (Fig. 1b, see A5), and (b) *Rhbdf1* KO mice using

CRISPR/Cas9-mediated gene editing (Fig. 1b, see A7), which results in alleles lacking the exon(s) with the translation initiation site ATG. We hypothesized that the mutant transcripts generated from the targeted KO-first or CRISPR/Cas9 alleles reinitiate transcription and translation in unexpected ways to rescue the more severe phenotype described by Christova et al. We sought to test our hypothesis by systematically characterizing the *Rhbdf1* single mutant mice and *Rhbdf1:Rhbdf2* double mutant mice generated in the present study, comparing the phenotypes with those reported in previously published studies (Fig. 1c).

Complete loss of exons in *Rhbdf1* (A5 allele) results in growth retardation, brain hemorrhage, and cardiac hypertrophy

To explore the function of RHBDF1 in vivo, we generated *Rhbdf1* homozygous-null (*Rhbdf1$^{-/-}$*) mice from heterozygotes with the A5 allele and validated *Rhbdf1* gene disruption by whole-exome sequencing (Exome-Seq, Fig. 2a) and polymerase chain reaction (PCR; Fig. 2b). Exome-Seq validated loss of exons 3 through 18, leading to deletion of almost all of the cytosolic N-terminus domain and all of the transmembrane domains. Filial crosses of heterozygotes resulted in near-Mendelian ratios, suggesting no embryonic lethality due to RHBDF1 deficiency. Moreover, all *Rhbdf1$^{-/-}$* newborns had a milk spot, and as neonates showed no gross deformities or malformations of major organs, except for growth retardation when compared with *Rhbdf1$^{+/-}$* littermates (Fig. 2c, left). The majority of *Rhbdf1$^{-/-}$* mice were found dead by P14, while the few remaining *Rhbdf1$^{-/-}$* neonates survived only to 3 to 4 weeks (Fig. 2d). These survivors appeared severely dehydrated and lethargic, and both female and male *Rhbdf1$^{-/-}$* survivors exhibited significantly lower body weight compared with *Rhbdf1$^{+/-}$* littermates (Fig. 2e).

In addition to the generalized growth deficiency of *Rhbdf1$^{-/-}$* survivors, specific abnormalities were observed in the brain and heart. Gross examination of whole brains from P21 *Rhbdf1$^{-/-}$* mice revealed hemorrhages, and histological examination of *Rhbdf1$^{-/-}$* brain sections confirmed several hemorrhages throughout the brain. These were evident in the ventricles adjacent to the brain stem, hippocampus, and hypothalamus (Fig. 2f), without significant inflammation. Additionally, P21 *Rhbdf1$^{-/-}$* mice displayed left atrial enlargement with left ventricular hypertrophy, which is reflected in the higher heart weight/body weight ratio in these homozygous mice compared to heterozygotes (Fig. 2g). Histological sections of the hearts revealed perivascular and interstitial fibrosis in the left ventricle and interventricular septum (Fig. 2h). In agreement with these data, heart-to-body-weight ratios revealed significant cardiac hypertrophy in the *Rhbdf1$^{-/-}$* animals. In contrast, no lesions were detected in hearts from *Rhbdf1$^{+/-}$* mice aged up to 1 year. Our findings in the present study—that *Rhbdf1$^{-/-}$* mice on a pure inbred C57BL/6N strain background have significantly lower body weight compared with heterozygous-null mice, with most *Rhbdf1$^{-/-}$* mice dying postnatally from brain hemorrhages and cardiac fibrosis—are similar in many ways to those of the Christova et al. study, but very different from those of Li et al. Notably, Christova et al. generated mice on two different strain backgrounds—C57BL/6J and 129S6/SvEvTac—and showed that whereas homozygotes on each background showed severe weight loss, brain hemorrhage, and cardiac infarction, and died by ~ 6 weeks, heterozygotes on each

Fig. 2 Complete deletion of exons in *Rhbdf1* results in growth retardation, brain hemorrhage, and cardiac hypertrophy (A5). **a** Exome-Seq showing loss of exons 2 through 18 in *Rhbdf1*$^{-/-}$ homozygous mutant mice generated using the definitive-null strategy. The red rectangle shows the exons lost in the mutant mice. **b** PCR of tail genomic DNA with all four primers shows that wildtype mice have one wildtype band (268-bp PCR product), heterozygous-null mice carry a wildtype band (268 bp) and a mutant band (546 bp), and homozygous-null mice carry one mutant band (546 bp). NTC, no template control. New England Biolabs 1 kb DNA Ladder was used as a molecular marker. Samples were run in duplicate. **c** Images of representative postnatal day 28 *Rhbdf1*$^{-/-}$ (left) and *Rhbdf1*$^{+/-}$ (right) male littermates. **d** Kaplan–Meier survival analyses showing that the lifespan of *Rhbdf1*$^{-/-}$ homozygous-null female (left panel) and male (right panel) mice is significantly shorter than that of heterozygous-null littermate controls. **e** Body weights of *Rhbdf1*$^{-/-}$ female ($n = 4$) and male ($n = 3$) mice compared with those of *Rhbdf1*$^{+/-}$ littermates ($n = 3$ for each sex) on postnatal day 21. Data represent mean ± S.D.; ***$p < 0.001$. **f** Brain hemorrhaging in *Rhbdf1*$^{-/-}$ mice. Upper left panel: multifocal hemorrhages (arrows) were observed in *Rhbdf1*$^{-/-}$ mice. All other panels: H&E-stained coronal sections of postnatal day 21 *Rhbdf1*$^{-/-}$ brain showing prominent hemorrhages in the brain stem, hippocampus, and hypothalamus. **g** Enlarged hearts in both female (left panel) and male (right panel) *Rhbdf1*$^{-/-}$ mice. Enlarged hearts were observed in pups older than 3 weeks of age. When we measured heart and body weights of *Rhbdf1*$^{-/-}$ ($n = 3$ females; $n = 4$ males) and *Rhbdf1*$^{+/-}$ ($n = 4$ females; $n = 5$ males) littermates at postnatal day 21, the heart-to-body-weight ratio (HW/BW) was significantly higher *Rhbdf1*$^{-/-}$ mice relative to *Rhbdf1*$^{+/-}$ littermates. Data represent mean ± S.D.; *$p < 0.05$, ***$p < 0.001$. **h** Cardiac complications in *Rhbdf1*$^{-/-}$ mice on postnatal day 21. Masson's trichrome staining of heart sections from a 3-week-old *Rhbdf1*$^{-/-}$ mouse showing interventricular septum and left ventricular fibrosis. Notably, both perivascular and interstitial fibrosis were observed in *Rhbdf1*$^{-/-}$ mice

background were normal. Collectively, these data suggest an essential role for RHBDF1 during postnatal development, and hereafter, the *definitive-null Rhbdf1* homozygous-null mice are referred to as *Rhbdf1*$^{null/null}$ mice.

Because the *Rhbdf1* and *Rhbdf2* genes have significant homology (~ 80%) in their sequences, we sought to address whether the phenotype of *Rhbdf1:Rhbdf2* double KO mice is more severe than that of *Rhbdf1*^{null} mice; i.e., embryonic or perinatal lethality.

Notably, $Rhbdf2^{-/-}$ mice are viable and fertile and do not display growth retardation or brain and heart defects [19–21]. Interestingly, in contrast to the viability of $Rhbdf1^{null/-null}$ embryos, the combined loss of $Rhbdf1$ and $Rhbdf2$ resulted in subviability. Of 64 live births from $Rhbdf1^{+/null}$ $Rhbdf2^{-/-}$ parents, there were four pups born that were homozygous-null for both genes ($Rhbdf1^{null/null}$ $Rhbdf2^{-/-}$) (Additional file 1: Figure S2a), suggesting subviability due to deficiency of both RHBDF1 and RHBDF2, and suggesting functional redundancy between RHBDF1 and RHBDF2 during embryonic development. Moreover, histopathological analysis of $Rhbdf1^{null/null}$ $Rhbdf2^{-/-}$ mice revealed an "eyelids open at birth" phenotype (Additional file 1: Figure S2b and S2c), suggestive of deficiencies in epidermal growth factor receptor (EGFR) signaling pathways [22, 23]. These findings—that the majority of $Rhbdf1^{null/null}$ $Rhbdf2^{-/-}$ mice die in utero—are similar to those of the study of Christova et al., who found that $Rhbdf1$:$Rhbdf2$ double KO mice on either the C57BL/6J or the 129S6/SvEvTac backgrounds result in embryonic lethality.

CRISPR/Cas9-mediated deletion of exons 2 and 3 in $Rhbdf1$ results in viable and fertile homozygous mutants with the A7 allele

We decided to investigate whether deletion of $Rhbdf1$ exons 2 and 3, which contain the translation initiation codon, rather than deletion of the entire protein-coding sequence of the target gene (as in the A5 allele), would result in mice that mimic $Rhbdf1^{null/null}$ mice (Fig. 2) or mice that exhibit an altered phenotype owing to creation of an N-terminally truncated abnormal protein product. We used CRISPR/Cas9 to generate this $Rhbdf1$ mutant allele (A7), which excised exons 2 and 3 by the NHEJ repair pathway (Fig. 3a and Additional file 1: Figure S3). Mutant mice that are homozygous for this allele are hereafter referred to as viable ($Rhbdf1^{v/v}$) mice. We observed that whereas heterozygous-viable ($Rhbdf1^{+/v}$) mice have one $wildtype$ band (1300 bp PCR product) and one mutant band (728 bp), homozygous-viable ($Rhbdf1^{v/v}$) mice have only the mutant band (728 bp) (Fig. 3b, left). The deletion of DNA sequence was further validated by Sanger's sequencing (Fig. 3b, right). None of the $Rhbdf1^{v/v}$ pups showed gross deformities or malformations of major organs in contrast to the majority of $Rhbdf1^{null/null}$ mice, which were found dead by P14 (Fig. 2). $Rhbdf1^{v/v}$ mice were healthy, viable, and fertile (Fig. 3c–e). Additionally, no specific abnormalities were observed in the brain, heart, kidney, spleen, skin, or liver by histological examination (Fig. 3f, g and Additional file 1: Figure S4a). This is also reflected in the normal heart weight/body weight ratio in $Rhbdf1^{+/v}$ and $Rhbdf1^{v/v}$ mice (Fig. 3h), compared with that of $Rhbdf1^{null/null}$ mice. These findings are similar in many ways to those of the Li et al. study, who found that the $Rhbdf1$ mutant mice they generated were indistinguishable from their littermate controls [18].

Exome-Seq of $Rhbdf1^{v/v}$ mice validated an in-frame 572-bp deletion in the $Rhbdf1$ gene, resulting in loss of exons 2 and 3 (Fig. 3i). To examine whether this deletion produces an alternative $Rhbdf1$ transcript, we performed RNA-Seq on RNA derived from spleens of $Rhbdf1^{v/v}$ mice, which showed the $Rhbdf1^{v/v}$ transcript lacked exons 2 and 3, but contained all of the other exons (1, and 4 through 18) (Fig. 3j). Because the normal translation initiation site (ATG) in exon 3 is missing, we predicted that the $Rhbdf1^{v/v}$ transcript could produce a protein variant or a truncated RHBDF1 protein by utilizing
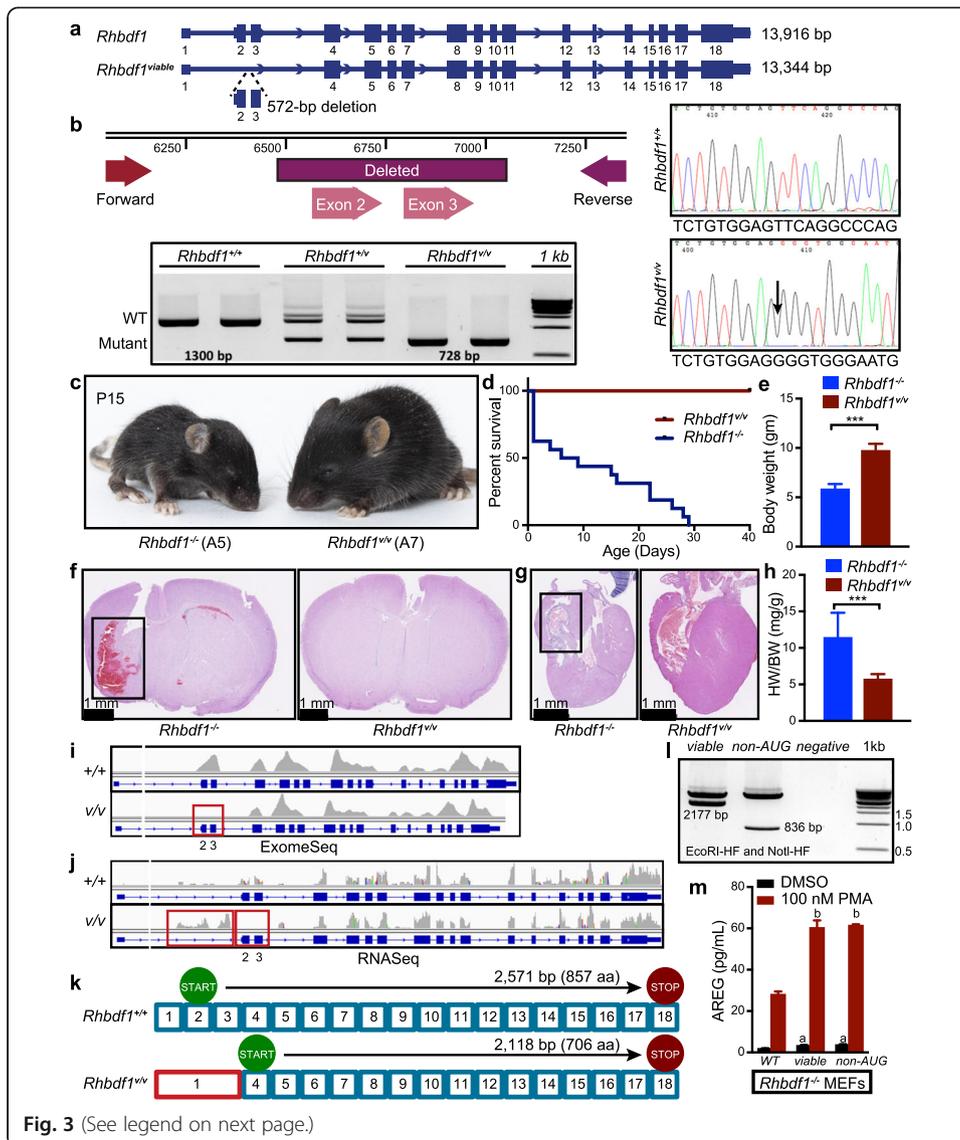
Hosur *et al. Genome Biology*   (2020) 21:168

Page 8 of 25

Fig. 3 (See legend on next page.)

Hosur et al. Genome Biology   (2020) 21:168

Page 9 of 25

(See figure on previous page.)

Fig. 3 Deletion of exons 2 and 3 in *Rhbdf1* results in viable and healthy mice (A7). **a** CRISPR/Cas9-mediated deletion of exons 2 and 3 resulted in an in-frame 572-bp deletion in the *Rhbdf1* gene (hereafter referred to as viable mice, *Rhbdf1*$^{v/v}$). **b** The schematic shows the deleted exons and the PCR strategy to amplify the region. PCR of genomic DNA with primers flanking exons 2 and 3 shows that wildtype mice have one wildtype band (1300-bp PCR product), heterozygous mice carry a wildtype band (1300 bp) and a mutant band (728 bp), and homozygous-null mice carry one mutant band (728 bp). Samples were run in duplicate. DNA sequencing further validated an in-frame 572-bp deletion in genomic DNA (right). Also, see Additional file 1: Figure S4. **c** Images of representative postnatal day 15 *Rhbdf1*$^{-/-}$ mice (left) and *Rhbdf1*$^{v/v}$ mice (right). **d** Kaplan–Meier survival analyses showing that the lifespan of *Rhbdf1*$^{v/v}$ mice (A7) is significantly longer than that of *Rhbdf1*$^{-/-}$ mice (A5). **e** Body weights of *Rhbdf1*$^{v/v}$ mice compared with those of *Rhbdf1*$^{-/-}$ mice on postnatal day 21. Data represent mean ± S.D.; ***$p < 0.001$. **f** H&E-stained coronal sections of a postnatal day 21 *Rhbdf1*$^{-/-}$ brain showing prominent hemorrhages in the brain; in contrast, there was no evidence of brain hemorrhaging in *Rhbdf1*$^{v/v}$ mice. **g** Cardiac abnormalities were observed in *Rhbdf1*$^{-/-}$ mice but not in *Rhbdf1*$^{v/v}$ mice (right). The black rectangle illustrates cardiac fibrosis in *Rhbdf1*$^{-/-}$ mice (left). **h** Also, heart weight was significantly lower in *Rhbdf1*$^{v/v}$ mice than in *Rhbdf1*$^{-/-}$ mice. Data represent mean ± S.D.; ***$p < 0.001$. **i** Exome-Seq showing loss of exons 2 and 3 in *Rhbdf1*$^{v/v}$ mutant mice and not in *Rhbdf1*$^{+/+}$ mice. The red rectangle shows the exons lost in the mutant mice. **j** RNA sequencing analysis of spleens isolated from *Rhbdf1*$^{v/v}$ mice indicating that CRISPR/Cas9-mediated deletion of exons 2 and 3 does not result in nonsense-mediated decay of mutant *Rhbdf1*$^{v/v}$ mRNA. Note the boxed region preceding the lost exons 2 and 3 indicates the presence of alternative transcripts. **k** Although the translation initiation site in exon 2 is deleted in *Rhbdf1*$^{v/v}$ mice, sequence analysis of *Rhbdf1*$^{v/v}$ cDNA revealed that the next in-frame translation initiation site (ATG) was in exon 4, which could potentially result in an N-terminally truncated protein to rescue the severe pathology of *Rhbdf1*$^{-/-}$ mice (A5). **l** Double digests of C-terminal Myc-DDK-tagged *Rhbdf1*^{viable} and *Rhbdf1*$^{non-AUG}$ vectors using EcoRI-HF and NotI-HF restriction enzymes. **m** Rescue experiments in *Rhbdf1*$^{-/-}$ MEFs. RHBDF1 deficiency suppresses stimulated secretion of AREG in primary MEFs (see Additional file 1: Figure S5); hence, we performed rescue experiments to restore stimulated AREG secretion in *Rhbdf1*$^{-/-}$ MEFs (A5); *Rhbdf1*$^{-/-}$ MEFs were transiently transfected with 2 μg of either a full-length vector, or a *Rhbdf1*^{viable} vector, or a *Rhbdf1*$^{non-AUG}$ vector, using Lipofectamine LTX. Forty-eight hours post-transfection, cells were stimulated overnight with either DMSO or 100 nM PMA, and cell culture supernatants were analyzed using a mouse AREG ELISA kit. Data represent mean ± S.D.; (a) significantly different from a full-length vector-transfected and DMSO-stimulated cells; (b) significantly different from a full-length vector-transfected and PMA-stimulated cells

the next in-frame translation initiation site in exon 4. Notably, *Rhbdf1* transcript variants X2, X3, and X4 all lack exons 2 and 3 [24], but generate an alternative mRNA transcript, which is likely to result in a 706-amino acid product utilizing the next in-frame ATG in exon 4 (Fig. 3k and Additional file 1: Figure S4b).

To determine whether the *Rhbdf1*$^{v/v}$ truncated protein is functional and whether it could rescue RHBDF1 deficiency, we performed in vitro stimulated secretion assays in primary mouse embryonic fibroblasts (MEFs). Previous studies have shown that RHBDF1 regulates protein kinase C activator phorbol ester (PMA)-mediated stimulated secretion of epidermal growth factor receptor (EGFR) ligands, including amphiregulin (AREG) [20]. Here, using PMA-induced stimulated secretion assays in *Rhbdf1*$^{+/+}$ and *Rhbdf1*$^{-/-}$ primary MEFs (A5 allele), we found that, upon stimulation with 100 nM PMA, *Rhbdf1*$^{+/+}$ MEFs showed a significant increase in AREG secretion compared with *Rhbdf1*$^{-/-}$ MEFs, indicating that RHBDF1 deficiency suppresses AREG secretion (Additional file 1: FigureS5). Hence, we asked whether the variant proteins can rescue the "stimulated secretion of AREG" phenotype, by transiently transfecting *Rhbdf1*$^{v/v}$-expressing vector (Fig. 3l) in *Rhbdf1*$^{-/-}$ primary MEFs, and measuring AREG levels following stimulation with PMA (Fig. 3m). Stimulation with PMA significantly increased AREG levels in *Rhbdf1*$^{v/v}$-transfected MEFs compared with empty vector-transfected MEFs, suggesting that *Rhbdf1*$^{v/v}$ is capable of inducing AREG secretion and thereby rescuing the phenotype.

Because a majority of double mutant *Rhbdf1^{null/null}^* *Rhbdf2^{−/−}^* mice die in utero, we asked whether the *Rhbdf1^{v/v}^* mutation can rescue the subviability of *Rhbdf1^{null/null}^* *Rhbdf2^{−/−}^* mice, by generating *Rhbdf1^{v/v}^* *Rhbdf2^{−/−}^* mice. In contrast to the subviability of *Rhbdf1^{null/null}^* *Rhbdf2^{−/−}^* mice, *Rhbdf1^{v/v}^* *Rhbdf2^{−/−}^* mice were viable, healthy, and fertile (Additional file 1: Figure S6a). Nevertheless, *Rhbdf1^{v/v}^* *Rhbdf2^{−/−}^* mice exhibited wavy-coated hair and "eyelids open at birth" phenotypes (Additional file 1: Figure S6b and S6c), indicative of alterations in epidermal growth factor receptor (EGFR) signaling pathways. Histopathological analysis revealed no evidence of heart or brain abnormalities (Additional file 1: Figure S6d).

Literature suggests that, in addition to the canonical AUG start codons, non-AUG start codons are capable of initiating translation [25–28]. Notably, *Rhbdf1* protein-coding transcript 206 [24] utilizes a non-AUG start codon, CGC (arginine, R), rather than AUG (methionine, M) (Additional file 1: Figure S7a). We next sought to examine whether mRNA translation initiation from non-AUG start codons can generate functional protein-coding transcripts that rescue the phenotype. Since there are no validated RHBDF1-specific antibodies, we obtained C-terminal Myc-DDK tagged expression vector for the non-AUG mutant transcript, and heterologously expressed them in 293T cells. Western blotting revealed that non-AUG vector, based on their molecular masses, generated expected size bands (~ 30 kDa) while transfection with an empty tagged vector did not produce any bands (Additional file 1: Figure S7b). Additionally, when the non-AUG-expressing vector (Fig. 3l) was transfected into *Rhbdf1^{−/−}^* primary MEFs, it induced significant AREG secretion (Fig. 3m). This result is similar to the *Rhbdf1^{v/v}^* transcript in transfected MEFs, suggesting that the non-AUG start codon CGC is capable of generating a functional protein-coding transcript. Thus, we conclude that mutant *Rhbdf1^{v/v}^* protein products rescue the severe phenotype observed in mice deficient in RHBDF1 (both Christova et al. and our *Rhbdf1^{null/null}^* mice), because of the use of alternative AUG and non-AUG start codons that nevertheless give rise to mostly functional proteins.

Targeted KO-first allele *Rhbdf1* homozygous mutants (A3) express truncated protein isoforms that rescue the pathology of *Rhbdf1^{null/null}^* mice

Additional crosses of mice carrying the *Rhbdf1* KO-first allele to Cre transgenic mice resulted in excision of exons 4–11 flanked by the *loxP* sites (Fig. 4a). These homozygous mutant mice were *generated by* Li et al. and are hereafter referred to as viable 2 mice (*Rhbdf1^{v2/v2}^*). Exome-Seq of these mice validated deletion of exons 4–11 in the *Rhbdf1* gene (Fig. 4b). RT-PCR (Fig. 4c *and* Additional file 1: Figure S8) and RNA-Seq (Fig. 4d) confirmed that exons 4 through 11 are missing in the transcript, while exons 12 through 18 remain intact. The structure of the *Rhbdf1^{v2/v2}^* transcript suggests that exons 3 and 12 might be spliced together. To test this, we performed RT-PCR on the *Rhbdf1^{v2/v2}^* transcript using probes in exon 3 and in exon 12. Instead of an expected 209-bp product suggesting that exons 3 and 12 are joined together, we detected a 324-bp PCR product, indicating insertion of a short 115-bp DNA fragment. Sanger sequencing of this product revealed that the *Rhbdf1^{v2/v2}^* mRNA retained the *En2* splice acceptor sequence (Additional file 1: Figure S8b,c) from the cassette (Additional file 1: Figure S8d). Nonetheless, the resultant *Rhbdf1^{v2/v2}^* mutant transcript retains an in-

Fig. 4 (See legend on next page.)
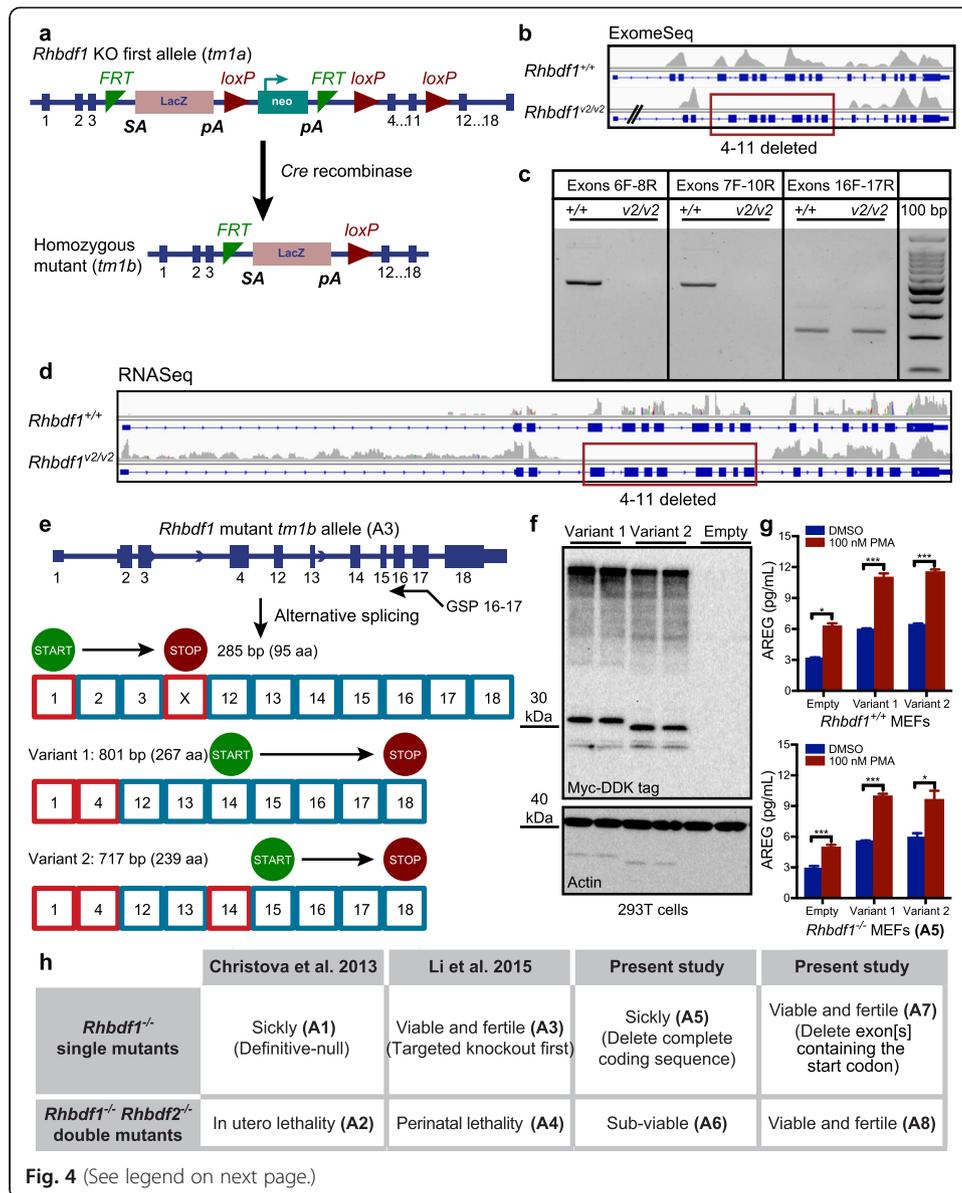
(See figure on previous page.)

Fig. 4 Targeted KO-first targeting strategy in *Rhbdf1* (A3) generates novel transcripts and N-terminally truncated functional proteins. **a** Schematic of the strategy used by Li et al. for generation of *Rhbdf1*$^{-/-}$ homozygous mutant mice; the *Rhbdf1* KO-first allele was crossed to Cre transgenic mice to excise the floxed gene segment (exons 4–11), generating *Rhbdf1*$^{-/-}$ homozygous mutant mice (hereafter referred as viable2 mice, *Rhbdf1*$^{v2/v2}$ mice). **b** Whole-exome sequencing of spleen tissue from *Rhbdf1*$^{v2/v2}$ mice showing loss of exons 4 through 11 in *Rhbdf1*$^{v2/v2}$ mutant mice. **c** RT-PCR on spleens from *Rhbdf1*$^{+/+}$ and *Rhbdf1*$^{v2/v2}$ mutant mice using primers to amplify exons 6 through 8, exons 7 through 10, and exons 16 and 17. Exons 4–11 are deleted in *Rhbdf1*$^{v2/v2}$ mutant mice; hence, no amplicons were generated using either exon 6 forward and exon 8 reverse, or exon 7 forward and exon 10 reverse, primers. However, exon 16 forward and exon 17 reverse primers generated a 211-bp product. **d** RNA-Seq analysis of spleens from *Rhbdf1*$^{v2/v2}$ mutant mice indicating loss of exons 4 through 11; however, there is strong evidence for mutant mRNA, as indicated by the presence of the rest of the transcript, which encodes exons 12 through 18 and is not degraded by the nonsense-mediated decay mechanism. **e** Schematic representation of exons and introns in the *Rhbdf1*$^{v2/v2}$ mutant allele. 5′ RACE using a gene-specific exon 16–17 fusion primer (GSP) was used to obtain 5′ ends of the *Rhbdf1*$^{v2/v2}$ mutant mRNA. We identified several novel mutant mRNAs with different translation initiation sites that could potentially generate N-terminally truncated RHBDF1 mutant proteins. See supplemental figures for variant protein and 5′ UTR sequences. Alternative exons are indicated as red boxes; predicted translation initiation sites are indicated by "START," and termination codons are indicated by "STOP." **f** C-terminal Myc-DDK-tagged *Rhbdf1*$^{v2/v2}$ variant protein 1 (lanes 1, 2) or variant protein 2 (lanes 3,4), or empty vector (lanes 5, 6) were transiently expressed in 293T cells, and cell lysates were analyzed using western blotting with FLAG-specific antibody. After visualization of blots with a G:Box chemiluminescent imaging system, blots were washed, blocked in 5% nonfat dry milk, and re-probed with anti-actin antibody. **g** Rescue of phenotype in *Rhbdf1*$^{-/-}$ MEFs. *Rhbdf1*$^{+/+}$ (top) and *Rhbdf1*$^{-/-}$ (bottom) MEFs were transiently transfected with 2 μg of either variant 1 or variant 2 vectors, or an empty vector, using Lipofectamine LTX. Forty-eight hours post-transfection, cells were stimulated overnight with either DMSO or 100 nM PMA, and cell culture supernatants were analyzed using a mouse AREG ELISA kit. Data represent mean ± S.D.; *$p < 0.05$, ***$p < 0.001$. **h** Comparison of phenotypes of previously generated *Rhbdf1* single mutant and *Rhbdf1:Rhbdf2* double mutant mice with those of mice generated in the present study. The viability of *Rhbdf1*$^{v/v}$, *Rhbdf1*$^{v2/v2}$, and *Rhbdf1*$^{v/v}$ *Rhbdf2*$^{-/-}$ double mutant mice demonstrates that the incomplete coding sequences not degraded by NMD have functional significance and, moreover, that targeted KO-first conditional mutagenesis (A3) or CRISPR/Cas9-mediated deletion of exon(s) (A7), in contrast to complete deletion of exons (A1 and A5), can yield gain-of-function alleles

frame termination codon (TAA), which will result in a 10.8-kDa short truncated protein product (Additional file 1: Figure S8e).

We next sought to explore whether the conditional *Rhbdf1*$^{v2/v2}$ mutation leads to additional truncated proteins via the use of an alternative promoter usage or via exon skipping to reinitiate translation [15, 29, 30]. To evaluate the 5′ ends of mutant *Rhbdf1*$^{v2/v2}$ mRNA transcripts, we performed 5′ RACE on first-strand cDNA using a gene-specific exon 16–17 fusion primer, based on the RT-PCR evidence for the presence of a 211-bp product (Fig. 4c). The resulting RACE products were characterized by Sanger sequencing following agarose gel electrophoresis and In-Fusion cloning. Sequence analyses of identified clones revealed the existence of an alternative exon 1, 199 bp upstream of the conventional exon 1 (Additional file 1: Figure S9 and S10). Interestingly, we found evidence for two potential transcripts that utilize this alternative exon 1 (Additional file 1: Figure S11 and S12) that could generate N-terminally truncated *Rhbdf1*$^{v2/v2}$ proteins by alternative splicing (Fig. 4e). We also obtained C-terminal Myc-DDK-tagged expression vectors for variant 1 and 2 mutant transcripts, and heterologously expressed them in 293T cells (Additional file 1: Figure S13). Western blotting revealed that both variants (1 and 2), based on their molecular masses, generated expected size bands (~ 32-kDa and ~ 29-kDa) while transfection with an empty tagged vector did not produce any bands (Fig. 4f).

Lastly, using PMA-induced stimulated secretion assays, we observed that, first, in both *Rhbdf1*$^{+/+}$ and *Rhbdf1*$^{-/-}$ MEFs, transfection alone with either variant protein 1 or

2 significantly increased AREG secretion compared with that of the empty vector (Fig. 4g). Second, stimulation with PMA further increased AREG levels in both variant protein 1- or 2-transfected MEFs compared with empty vector transfected MEFs, suggesting that both of the N-terminally truncated RHBDF1 variant proteins are capable of inducing AREG secretion and thereby rescuing the phenotype (Fig. 4g). Collectively, these data suggest that the lack of spontaneous pathological phenotypes in the $Rhbdf1^{v2/v2}$ mutant mice could be explained by the presence of residual mRNA transcripts and protein isoforms that can prevent the more severe phenotype observed in the $Rhbdf1^{null/null}$ mice, and that the $Rhbdf1^{v2/v2}$ mice with the targeted KO-first allele (A3) are not homozygous-null (Fig. 4h).

Identifying unexpected translation by analyzing incomplete protein-coding sequences (CDS) evading nonsense-mediated decay

We next investigated whether we could find mouse and human protein-coding genes that are likely predisposed to targeted KO-first- and CRISPR/Cas9-mediated illegitimate or unexpected transcription and translation. Several factors, including exon skipping, alternative promoter usage, and translation reinitiation at AUG and at non-AUG start codons, could result in abnormal phenotypes. Identifying mouse and human protein-coding genes that express incomplete coding transcripts (Additional file 2: Table S1, Additional file 3: Table S2, Additional file 4: Table S3, Additional file 5: Table S5) that are not subjected to NMD could help identify genes resilient to targeted KO-first or frameshift indels. We found that at least 35% of mouse and 45% of human protein-coding genes generate protein-coding transcripts that are not eliminated by NMD (Fig. 5a). For instance, the Ensembl site lists nine transcripts (*Rhbdf1*-201 through *Rhbdf1*-209) for mouse *Rhbdf1* (Fig. 5b). Of the four protein-coding *Rhbdf1* transcripts (201, 206, 207, and 209), transcript 206 has a 5′ incomplete CDS and is comprised of only exons 14 through 18. As such, this transcript is not likely to be degraded by NMD when alleles with targeting event 5′ of these exons are generated. Indeed, our tests in the present study were focused on amplifying exons 15 through 18 in the $Rhdbf1^{v2/v2}$ allele (Fig. 4c). Additionally, 5′ RACE on first-strand $Rhdbf1^{v/v}$ and $Rhdbf1^{v2/v2}$ cDNA was performed using a gene-specific exon 16–17 fusion primer (Fig. 4e). Importantly, targeted KO-first mutagenesis excised only exons 4 through 11, leaving the exons in the *Rhbdf1*-206 transcript (14 to 18) untouched and thus not likely to be subject to NMD.

BAZ2A is another example of a gene that was recently identified during systematic characterization of frameshift KO mutations in HAP1 cells showing residual protein expression [15], that we could identify here based on the presence of incomplete CDS (Fig. 5c). In Additional file 6: Table S5, we present additional examples of the genes that retain residual protein expression/function upon CRISPR/Cas9-induced indels in recent literature [14–17, 28–31] and that we validate as good candidates for unanticipated translation based on the incomplete CDS approach.

We also analyzed 3674 conditional-ready mouse strains [32] using the incomplete CDS approach (Additional file 7: Table S6), and determined that at least 40% of conditional-ready alleles can skew the mouse phenotype (Fig. 5d). For instance, the *Bbs5* conditional-ready allele is likely to generate a 5′ incomplete CDS and an N-
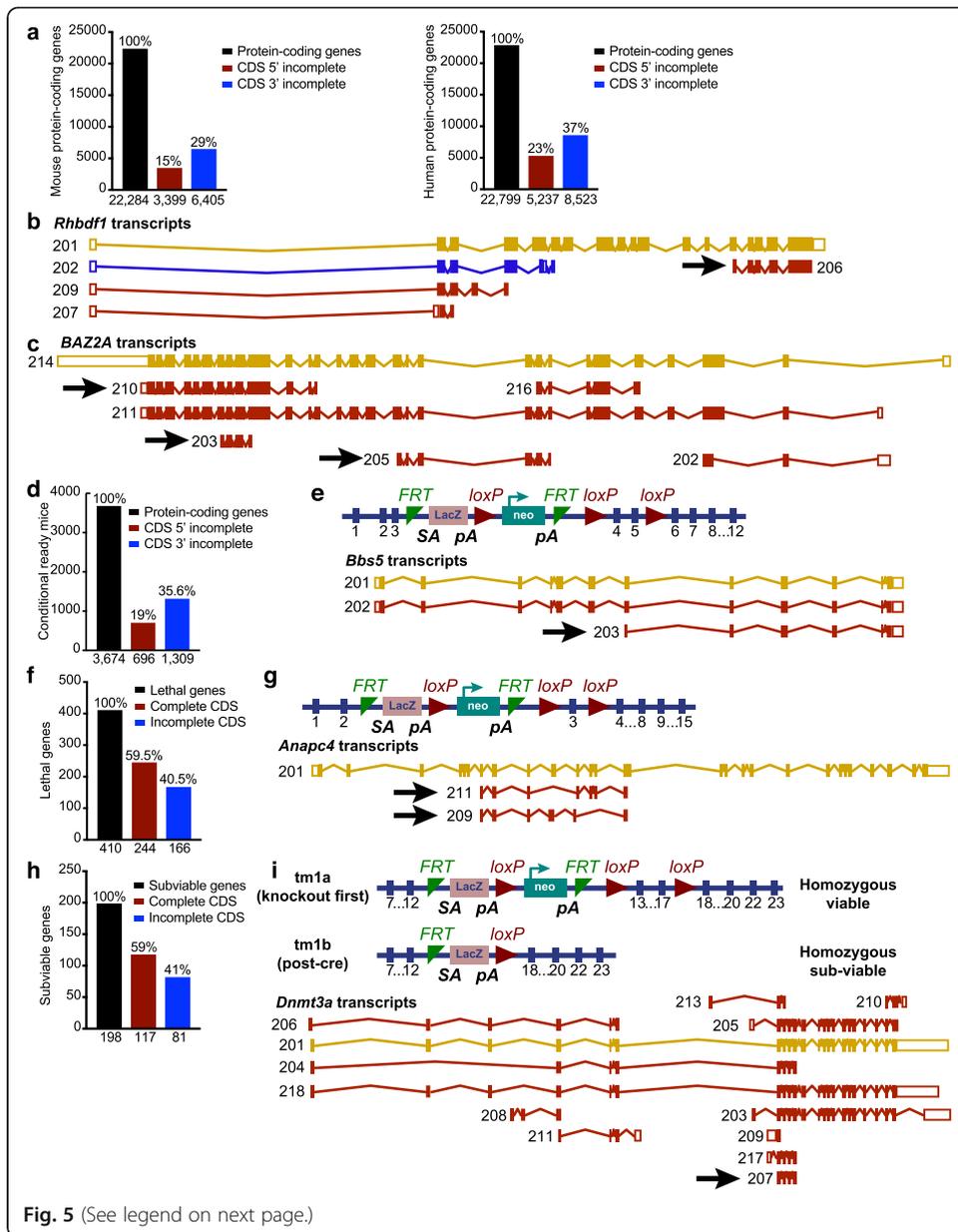
Fig. 5 (See legend on next page.)

Hosur *et al. Genome Biology*   (2020) 21:168

Page 15 of 25

(See figure on previous page.)
Fig. 5 An "incomplete protein-coding CDS" approach to identifying on-target unintended transcription and translation reinitiation. **a** Bioinformatics analyses of mouse (left) and human (right) protein-coding genes with incomplete 5′ CDSs and incomplete 3′ CDSs reveals that ~ 50% of genes are potentially capable of generating shorter transcripts and truncated proteins. The physiological significance of these incomplete CDSs is unclear. However, it is yet to be determined whether gene-targeting strategies must be designed to ensure the deletion of these incomplete, yet existing coding sequences that are not degraded by NMD. Notably, a 3′ incomplete CDS is a protein-coding transcript which is missing the stop codon due to incomplete evidence and a 5′ incomplete CDS is a protein-coding transcript which is missing the start codon due to incomplete evidence. **b** The various transcripts of the mouse *Rhbdf1* gene listed in Ensembl. In this scenario, the definitive-null design, rather than either the targeted KO-first or CRISPR/Cas9-mediated deletion of exon(s) containing the start codon, would prevent expression of a short 5′ incomplete *Rhbdf1-*206 transcript. **c** CRISPR/Cas9-mediated frameshift mutation failed to induce complete deficiency of *BAZ2A* in HAP1 cells; however, based on the presence of 5′ and 3′ incomplete CDSs (arrows), *BAZ2A* is likely to generate truncated proteins upon targeted KO-first- or CRISPR/Cas9-induced on-target mutagenesis. **d** Bioinformatics analyses of 3674 conditional-ready mouse strains identified that nearly 55% of conditional-ready genes generate incomplete 5′ and 3′ CDSs. **e** Schematic of *Bbs5* KO-first allele in which the "lacZ reporter gene and neomycin resistance gene" flanking FRT recognition sites is placed downstream of the third exon of the gene, and exons 4 and 5 are floxed by *loxP* sites. Following Cre-mediated recombination, despite missing exons 4 and 5, *Bbs5* 5′ incomplete CDS 203 may generate an N-terminally truncated protein. **f** Bioinformatics analyses of 410 lethal IMPC mouse strains identified that 166 essential genes that generate incomplete 5′ and 3′ CDSs. **g** Schematic of *Anapc4* KO-first allele in which the "lacZ reporter gene and neomycin resistance gene" flanking the FRT recognition sites is placed downstream of the second exon of the gene, and exon 3 is floxed by *loxP* sites. Although Cre recombination deletes exon 3, translation reinitiation from incomplete CDSs 209 and 211 may rescue the lethal phenotype of *Anapc4* homozygous-null mice. **h** Analyses of 198 subviable IMPC mouse strains identified 81 genes generating incomplete 5′ and 3′ CDSs. **i** The targeted KO-first strategy in *Dnmt3a* results in conflicting phenotypes. Whereas the *tm1a* KO-first allele generates homozygous-viable mice, the *tm1b* allele generates homozygous subviable mice (top). Multiple protein-coding transcripts of *Dnmt3a* are shown. The *Dnmt3a* 207 protein-coding transcript has incomplete 5′ and 3′ CDSs (arrow) (bottom)

terminally truncated protein despite missing exons 4 and 5, which are lost upon Cre-mediated recombination (Fig. 5e).

In contrast to mutation of gene function via unsolicited translation following CRISPR/Cas9 or targeted KO-first gene targeting, can unsolicited translation be advantageous for exploring the function of genes that are essential for survival? The International Mouse Phenotyping Consortium (IMPC) systematically characterized 5000 KO mouse strains and identified 410 lethal and 198 subviable genes [33]. Here, we analyzed all of the lethal and subviable genes (Additional file 7: Table S6) and found that, upon Cre recombination of conditional-ready alleles, nearly 40% of the lethal genes and 40% of the subviable genes may result in unwanted translation and could potentially partially rescue the mouse phenotype generated by a null mutation (Fig. 5f, g). We show an example of the *Anapc4* KO-first allele, in which the "lacZ reporter gene and neomycin resistance gene" flanking the FRT recognition sites for FLP recombinase is placed downstream of the second exon of the gene, and exon 3 is floxed by *loxP* sites for recognition by Cre recombinase. Flp recombinase removes the lacZ and neomycin cassette, allowing reversal to a conditional-ready allele, while Cre recombinase deletes exon 3, presumably generating null or conditional mutants (Fig. 5h). However, *Anapc4* expresses incomplete protein-coding transcripts that lack exon 3 and that evade NMD (*Anapc4*–209 and – 211), suggesting that these protein-coding transcripts could rescue the phenotype and impact interpretation of the phenotype.

In accordance with our observations, emerging evidence from IMPC suggests that several mouse strains generated using the targeted KO-first strategy exhibit conflicting phenotypes: *Dnmt3a* (*tm1a* allele is homozygous viable; *tm1b* allele is homozygous

subviable) (Fig. 5i); *Foxj3* (*tm1a* is homozygous viable; *tm1b* is homozygous lethal); and *Slc20a2* (*tm1a* is homozygous viable; *tm1b* is homozygous subviable), suggesting at least partial rescue of gene function due to residual protein activity (Table 1 and Additional file 8: Table S7). Notably, both the *tm1a* and *tm1b* mouse strains, KO-first and post-Cre, respectively, should exhibit the same phenotype. Thus, we strongly believe that, despite rigorous quality control of mutant mouse strains by the IMPC, additional post-genomic-DNA characterization is required to ensure elimination of any residual proteins of the targeted allele that might interfere with the phenotype. This is particularly important when novel sequences are incorporated, such as those introduced by targeted KO-first, making it extremely difficult to predict any novel mRNA transcripts that may be generated.

Discussion

Several recent studies have reported that CRISPR/Cas9-mediated on-target gene KO can produce novel unintended mRNA isoforms and truncated proteins that could at least partially rescue activity of the targeted gene; Lalonde and colleagues found that despite CRISPR/Cas9-induced on-target frameshift mutations in the phosphatase and actin regulator 1 *(PHACTR1)* gene in human aortic endothelial cells, exon skipping results in novel functional *PHACTR1* transcripts [29]. Also, in zebrafish lines, CRISPR/Cas9-generated KOs display irregular phenotypes, owing to activation of cryptic splice sites in mutant mRNA transcripts [17]. Similarly, to examine whether CRISPR/Cas9 mutations induce exon skipping to generate in-frame transcripts and truncated proteins with functional outcomes, Mou et al. targeted exon 3 of the beta-catenin gene *Ctnnb1*

Table 1 The KO-first gene disruption strategy results in conflicting phenotypes in homozygous mutant mice

Mice generated using KO-first strategy	Homozygous viable	Homozygous subviable	Homozygous lethal
*Bbox1**	em1	tm1b	
Caskin1	em1	tm1b	
*Cpsf3**		tm1b	tm1a
*Dbn1**	em1	tm1b	tm1b
*Dnmt3a**	tm1a	tm1b	
Dppa1	tm1a	tm1b	
Exoc3l2	tm1a		tm1b
*Foxj3**	tm1a		tm1b
Ino80		tm1b	tm1a
*Itm2c**	em2	tm1b	
*Kdm8**		tm1a	tm1b
*Lmna**		tm1a	tm1b
*Nxn**		tm1a	tm1b
*Rundc1**		tm1b	tm1a
*Slc20a2**	tm1a	tm1b	
*Tox3**	em1	tm1b	
Zranb1	tm1a	tm1b	

Analysis of 427 conditional-ready mouse strains generated using the KO-first strategy reveals that at least 4% of mouse strains show conflicting phenotypes. *Notably, 12 of 17 lethal/subviable genes express incomplete protein-coding sequences that could generate truncated proteins to potentially skew the phenotype of homozygous mutant mice

in a cell line, which induced indels as expected. However, unexpected splicing together of exons 2 and 4, or of exons 2 and 5, resulted in novel protein isoforms [30]. Immunoblotting with beta-catenin specific antibody revealed the presence of mutant proteins lacking exon 3-encoded amino acid residues, suggesting that indels induced by CRISPR/Cas9-sgRNAs can lead to skipping of edited exons, forcing production of truncated proteins. Makino et al. also found that biallelic on-target frameshift mutations in the zinc finger protein-encoding gene *Gli3* did not result in homozygous-null alleles, and owing to illegitimate translation, GLI3 protein expression persisted in mouse NIH3T3 cells [14]. More recently, Kapahnke et al. reported that editing of the flotillin-1 gene (*FLOT1*) in HeLa cells resulted in translation of an aberrant protein with a dominant negative effect due to alternative splicing [31]. A more comprehensive analysis of residual protein expression in HAP1 cells following CRISPR/Cas9-induced frameshift mutations by Smits and colleagues further supports the observation that truncated proteins can be generated by frameshift mutations by either skipping of the edited exon or translation reinitiation [15]. Nevertheless, there have been no reports demonstrating the physiological significance of truncated protein isoforms generated by mutant mRNA transcripts to date.

Given the possibility that conflicting data with respect to phenotype can result from differences in the outcomes of different gene-targeting methods [1, 18], we comprehensively characterized *Rhbdf1* mutant alleles generated by distinct gene-targeting strategies—definitive-null, CRISPR/Cas9, and targeted KO-first. Initiating with Exome-Seq, our workflow involved RT-PCR, RNA-Seq, 5′ RACE-PCR, heterologous expression of tagged expression vectors in 293T cells, and functional rescue assays. We find that whereas *complete loss of exons in Rhbdf1$^{null/null}$* mice causes growth retardation, brain and heart defects, and death of the majority of mice by P14, CRISPR/Cas9 *Rhbdf1$^{v/v}$* mice, lacking the start codon, and targeted KO-first *Rhbdf1$^{v2/v2}$* mice, lacking exons 4 through 11, do not display growth retardation or any other pathologies observed in *Rhbdf1$^{null/null}$* mice. Additionally, based on our previous findings showing that N-terminally truncated RHBDF2 causes a gain-of-function phenotype [20], and our data here showing that the *Rhbdf1$^{v/v}$* allele rescues the subviability of *Rhbdf1$^{null/null}$* *Rhbdf2$^{-/-}$* mice, causing a wavy hair coat phenotype and inducing enhanced stimulated secretion of AREG in both *Rhbdf1$^{+/+}$* and *Rhbdf1$^{-/-}$* MEFs, we suggest that *Rhbdf1$^{v/v}$* and *Rhbdf1$^{v2/v2}$* mutant mice generate residual mRNA transcripts and protein isoforms and are gain-of-function mutations. These data also indicate that targeted KO-first alleles can generate gain-of-function alleles that rescue the severe phenotype caused by the null alleles (Additional file 1: Figure S14).

Moreover, our observation that genes can skip exons with stop codons to reinitiate transcription in vivo to potentially rescue biological activity is in accordance with previous in vitro studies [14, 15, 29, 30], suggesting the need for deeper analyses of genetically modified alleles. We propose that in addition to alternative mRNA splicing, other possible mechanisms that must be accounted for while characterizing mutant mouse strains include alternative promoter usage and translation initiation from non-AUG start codons [28, 34–36], both of which can often be assessed by 5′ RACE-PCR, residual protein analysis by immunoblotting, and functional rescue assays.

Systematic generation and phenotyping of thousands of KO mouse strains by the IMPC is one of the most significant biological resources available to researchers for

understanding mammalian gene function [37]. Despite advances in CRISPR-Cas genome editing technologies, generating conditional KOs is rather challenging [38], and furthermore, targeted KO-first alleles offer a unique advantage by facilitating monitoring of gene function in a tissue- and temporal-specific manner. Nevertheless, the present study emphasizes that nature can circumvent gene-targeting strategies to affect the phenotype of homozygous mutant mice. However, careful selection of gene-targeted ES clones and subsequent residual mRNA and protein analysis can considerably improve the likelihood of generating mice harboring conditional or loss-of-function alleles with pertinent phenotypes. Moreover, our data demonstrating that exon skipping in $Rhbdf1^{v/v}$ mice prevents the multi-organ pathology observed in RHBDF1-deficient mice and rescues embryonic lethality in RHBDF1:RHBDF2 double deficiency, indicates that the limitation of targeted KO-first and CRISPR/Cas9 approaches may provide opportunities to generate gain-of-function alleles that can either provide a partial function of a lethal gene to enable further examination of its phenotype in detail, or that can potentially be harnessed for therapeutic benefits in humans.

Methods

Mice

B6N(Cg)-$Rhbdf1^{tm1.1(KOMP)Vlcg}$/J mice, stock number 026424, were obtained from The Jackson Laboratory (JAX) Repository and were bred and maintained in a barrier facility at JAX with a 12-h light to 12-h dark cycle. Intercrossing heterozygotes generated $Rhbdf1^{-/-}$ mice. For generating the viable mice ($Rhbdf1^{v/v}$), fertilized oocytes from C57BL/6J mice (stock #664) were microinjected using two sgRNAs targeting intron 1–2 and intron 3–4, respectively. Each sgRNA was delivered at a concentration of 50 ng/μl, along with Cas9 mRNA (Trilink) at 100 ng/μl and Cas9 protein (PNABio) at 30 ng/μl. The microinjection reagents were combined in nuclease-free T10E0.1 pH 7.5 (Integrated DNA Technologies; Tris 10 mM, EDTA 0.1 mM) supplemented with Rnasin (Promega) at 0.2 U/μl. The sgRNAs were both designed as TRU guides [39] with 18 nt of target-specific sequence and utilized the backbone modifications described by Chen et al. [40]. The two guide target sites are shown here with PAM in parentheses: 5′-GGAGGATCTGTGGAGTTC(agg) and 5′-GAATCCCTGTTAGGTACC(agg). A total of 128 embryos were microinjected and transferred into 6 pseudopregnant females. One transfer failed to deliver any liveborn pups, but 31 mice (16F/13M) were born from the 107 embryos that were successfully transferred (29% survival). Nine of 31 (29%) pups showed a dropout allele (large deletion of the intervening sequence between the two cut sites). Genomic DNA from tails was amplified using PCR and then sequenced confirmed by Sanger sequencing. To validate germline transmission, mice harboring a 569-bp dropout were backcrossed to C57BL/6J mice and the resulting N1 offspring were intercrossed to generate the $Rhbdf1^{v/v}$ homozygous mutant mice. Intercrossing $Rhbdf1^{+/-}$ with $Rhbdf2^{-/-}$ mice generated $Rhbdf1^{-/-}Rhbdf2^{-/-}$ double homozygous-null pups, and intercrossing $Rhbdf1^{v/v}$ with $Rhbdf2^{-/-}$ mice generated $Rhbdf1^{v/v}Rhbdf2^{-/-}$ double homozygous-null mice. Generation of $Rhbdf2^{-/-}$ mice has been described previously [20]. Mice were provided food pellets (Purina LabDiet 5K52; 6% fat, autoclaved) and acidified water ad libitum.

Hosur *et al. Genome Biology*　　　(2020) 21:168

Page 19 of 25

Histology

Mice euthanized by CO_2 asphyxiation were gently perfused through intracardiac injection of PBS followed by injection of 4% paraformaldehyde. Tissues were fixed overnight by immersion in 10% neutral buffered formalin and then transferred to 70% ethanol. Tissue sections were stained with either hematoxylin and eosin (H&E) or Masson's trichrome stain.

5′ RACE

The SMARTer RACE 5′/3′ Kit was used for performing 5′-rapid amplification of cDNA ends (RACE) using total RNA according to the manufacturer's instructions. The integrity of total RNA isolated from spleen tissues was assessed using the Eppendorf BioSpectrometer. Total RNA (1 μg) was converted into RACE-Ready first-strand cDNA and was diluted with 10 μl of Tricine-EDTA. A 28-bp gene-specific primer (see Additional file 1: Figure S10) was used in touchdown PCR with 2.5 μl of 5′-RACE-Ready cDNA as template to generate the full-length cDNA. RACE products were analyzed following agarose gel electrophoresis, gel extraction, In-Fusion cloning, and Sanger sequencing.

Cell culture

Primary mouse embryonic fibroblasts (MEFs) were isolated from embryos at 13.5 dpc (days post coitus) as previously described [20] using Liberase DL (Sigma-Aldrich). Cells were grown in a humidified chamber at 37 °C with 5% CO_2. MEFs and 293T cells were grown in DMEM 10% fetal bovine serum and antibiotic/antimycotic (Thermo Fisher Scientific). MEFs were plated in 6-well dishes for transient transfection, western blotting, and ELISA. 293T cells obtained from the Viral Vector Core service at JAX were mycoplasma and MAP tested.

Site-directed mutagenesis and DNA cloning

Mouse expression plasmids containing cDNA encoding *Rhbdf1* was obtained from OriGene. The Q5® Site-Directed Mutagenesis Kit was used to introduce site-specific mutations in the mouse *Rhbdf1* plasmid according to the manufacturer's instructions and confirmed by Sanger sequencing. Primers were designed using the NEBaseChanger tool for site-directed substitutions and deletions, and purchased from IDT.

Genotyping

Two-millimeter tail tips were incubated overnight in lysis buffer containing Proteinase K at 55 °C in a shaker incubator and denatured at 95 °C for 5 min. One microliter of undiluted DNA was used for 25 μl PCR reaction. For genotyping primers and protocols, see Additional file 1: Figure S15.

Reverse transcriptase PCR (RT-PCR)

The Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) was used for cDNA synthesis from total RNA (1 μg) using oligo(dT)$_{18}$ primers. Reaction tubes containing either no reverse transcriptase or no template served as negative controls. cDNA synthesis was performed according to the manufacturer's instructions.

SYBR Green real-time PCR

PowerUp SYBR Green Master Mix was used for running the real-time PCR (Thermo Fisher Scientific). For a 10 µl/well PCR reaction, 100 ng of cDNA, 800 nM exon 16 forward primer (CCTCCTGCCTTTCCTCAATC), 800 nM exon 17 reverse primer (GAAGATGGCACTGGCTAGATT), and 5 µl of 2X master mix were mixed together and PCR was run on a ViiA 7 Real-Time PCR System using the standard cycling mode (primer $T_m \geq 60\,°C$).

Immunoblotting

293T cells seeded in six-well dishes and transiently transfected with expression plasmids were treated with 300 µl of RIPA lysis buffer (Cell Signaling Technology) and incubated on ice for 30 min. Cells were then centrifuged at 13,000×*g* at 4 °C for 10 min, 20 µg of total cell lysate was loaded in 4–20% Tris-Glycine Mini Gels, and samples were allowed to migrate for 90 min. Proteins were transferred onto a PVDF membrane using the iBlot™ Gel Transfer device, and membranes were blocked with 5% milk for 1 h at room temperature and exposed to either FLAG-specific (Origene, #TA50011) or actin (Cell Signaling Technology, #4970) antibodies overnight at 4 °C. Membranes were then washed and exposed to anti-mouse and anti-rabbit secondary antibodies, respectively, before visualizing the blots using a G:BOX gel documentation system (Syngene, Frederick, USA).

DNA isolation and exome sequencing

DNA was isolated from spleen using the Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's protocols. DNA quality was assessed using a Genomic DNA Screen Tape (Agilent Technologies) and Nanodrop 2000 spectrophotometer (Thermo Scientific). DNA concentration was assessed using a Qubit dsDNA BR Assay Kit (Thermo Scientific). The library was prepared by the Genome Technologies core facility at JAX using Kapa Hyper Prep (Roche Sequencing and Life Science) and SureSelectXT Mouse All Exon V2 Target Enrichment System (Agilent Technologies), according to the manufacturer's instructions. Briefly, the protocol entails shearing the DNA using the Covaris E220 Focused-ultrasonicator (Covaris), ligating Illumina-specific adapters, and PCR amplification. The amplified DNA library is then hybridized to Mouse All Exon probes, amplified using indexed primers, and checked for quality and concentration using High Sensitivity D5000 Screen Tape (Agilent Technologies) and quantitative PCR (KAPA Biosystems), according to the manufacturers' instructions. The library was sequenced 100-bp paired-end on the HiSeq 4000 (Illumina) using HiSeq 3000/4000 SBS Kit reagents (Illumina). Sequenced reads were filtered and trimmed for quality scores > 30 using a custom python script. Filtered Exome-Seq reads were aligned to *Mus musculus* GRCm38 using BWA-mem (0.7.9a-r786). Bam files were processed as needed (e.g., sorting, deduplication) with Picard/GATK (3.4–0).

RNA isolation and sequencing

RNA was isolated from spleen using the MagMAX mirVana Total RNA Isolation Kit (Thermo Fisher) and the KingFisher Flex purification system (Thermo Fisher). The tissue was lysed and homogenized in TRIzol Reagent (Thermo Fisher). After the addition

of chloroform, the RNA-containing aqueous layer was removed for RNA isolation according to the manufacturer's protocol, beginning with the RNA bead binding step. RNA concentration and quality were assessed using the Nanodrop 2000 spectrophotometer (Thermo Scientific) and the RNA Total RNA Nano assay (Agilent Technologies). The library was prepared by the Genome Technologies core facility at JAX using the KAPA mRNA HyperPrep Kit (KAPA Biosystems), according to the manufacturer's instructions. Briefly, the protocol entails isolation of polyA containing mRNA using oligo-dT magnetic beads, RNA fragmentation, first- and second-strand cDNA synthesis, ligation of Illumina-specific adapters containing a unique barcode sequence for each library, and PCR amplification. The library was checked for quality and concentration using the D5000 Screen Tape (Agilent Technologies) and quantitative PCR (KAPA Biosystems), according to the manufacturers' instructions. The library was sequenced 100 bp paired-end on the HiSeq 4000 (Illumina) using HiSeq 3000/4000 SBS Kit reagents (Illumina). Filtered RNA-Seq reads were aligned to *Mus musculus* GRCm38 using Bowtie2 (v2.2.0).

AREG ELISA

MEFs cultured in collagen-coated six-well plates were transiently transfected with indicated expression vectors and stimulated overnight with 100 nM PMA (R&D Systems, Minneapolis, USA). Cell culture supernatant was centrifuged at 2500×*g* at 4 °C for 5 min and 100 μl of supernatant was used to measure AREG protein levels by a Mouse Amphiregulin DuoSet ELISA Developmental Kit (#DY989, R&DSystems).

Incomplete human and mouse protein-coding sequences

Data were accessed from Ensembl via the SQL databases (useastdb.ensembl.org; user anonymous) [24]. The databases homo_sapiens_core_99_38 and mus_musculus_core_99_38 were queried for all transcripts where the attribute type was cds_start_NF(CDS start not found) or cds_end_NF (CDS end not found). Transcript IDs were used to select for gene ids, Ensembl stable gene ids, and biotypes assigned to transcripts and genes. Ensembl gene ids were used to select gene symbols via Ensembl Biomart. Data were collected in Microsoft Excel, and functions of Excel were used to combine (VLOOKUP) and filter the datatypes. We have developed a web application (https://incompletecds.jax.org) to identify protein-coding genes that could be predisposed to targeted KO-first- and CRISPR/Cas9-mediated unexpected transcription and translation.

Statistical analysis

Prism v7 software (GraphPad) was used to generate the Kaplan–Meier survival curves. Student's *t* test was used to determine whether there was a statistically significant difference between two groups. $P < 0.05$ was considered to be statistically significant.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13059-020-02086-0.

Additional file 1: Figure S1. Gene-targeting strategies for generating null or conditional mutant alleles: a. In the definitive-null design, the complete coding sequence of a gene is flanked by two *loxP* sites, and recombination with a cre recombinase excises the floxed gene segment, creating LacZ-tagged null alleles. Loss of the complete

---

coding sequence of the target gene ensures homozygous-null alleles. b. Targeted KO-first alleles harbor *FRT* sites flanking "*LacZ* and a *neo* gene," and also harbor *loxP* sites flanking either an exon (exon 2, left) or specific exons chosen to be excised (exons 4 through 11, right, in this case), permitting recombination with Flp or cre recombinases and enabling conditional mutagenesis. Application of cre recombinase deletes exons flanked by two *loxP* sites, generating LacZ-tagged null alleles (*tm1b*). Application of Flp recombinase to this allele is described in the legend for 1C. c. Application of Flp recombinase to targeted KO-first alleles reverts the knockout-first alleles into conditional-ready alleles by excising *lacZ* and a *neo* gene. Subsequent recombination of conditional-ready alleles with cre recombinase generates conditional mutant alleles (*tm1d*). **Figure S2.** Sub-viability in *Rhbdf1^-/-^ Rhbdf2^-/-^* double mutant mice (A6): a. Breeding strategy and the number of mice generated and genotyped by PCR b & c. *Rhbdf1^-/-^Rhbdf2^-/-^* double mutant mice displayed "eyelids open at birth" phenotype (arrows) in mouse embryos harvested on embryonic day 15 (e15.5) (B) and in pups at day 2 (C). No defects were observed in *Rhbdf1^+/-^Rhbdf2^-/-^* littermates. **Figure S3.** PCR and genotyping of *Rhbdf1^v/v^* allele: Schematic showing the TRU-guides (sgRNAs 2256 and 2257) used to excise exons 2 and 3 in the *Rhbdf1* gene. Also shown are the PCR primers and *Taq* DNA Polymerase thermocycling conditions used to genotype *Rhbdf1^v/v^* mice (A7). PCR products were sequenced to verify the 569-bp deletion. **Figure S4.** Histology and 5' RACE products of *Rhbdf1^v/v^* allele: a. H&E sections of *Rhbdf1^v/v^* mice displaying no abnormalities in the heart, brain, truncal skin sections, kidney, spleen, and liver tissues. b. Schematic representation of *Rhbdf1^v/v^* transcript variant with a different transcription start site. 5' RACE identified an alternative exon (highlighted in light orange) that splices with exon 4 to reinitiate transcription from a different start site in exon 4, as opposed to using the conventional start site in exon 2. Notably, splicing of alternative exon with exon 4 is also observed in other *Rhbdf1* predicted transcript variants X1, X2, X3, X4, X9, and X14. **Figure S5.** Stimulated secretion of AREG: Stimulated secretion of AREG in *Rhbdf1^+/+^* and *Rhbdf1^-/-^* primary MEFs, passage one, after overnight stimulation with 100 nM PMA. Following stimulation, cell-culture supernatants were analyzed using a mouse AREG ELISA kit. Data represent mean ± S.D; ***$p<0.001$. **Figure S6.** CRISPR/Cas9 *Rhbdf1^v/v^* allele (A7) reverses sub-viability in *Rhbdf1^-/-^ Rhbdf2^-/-^* double mutant mice: a. Images of representative postnatal day 21 *Rhbdf1^v/+^ Rhbdf2^-/-^* mice (left) and *Rhbdf1^v/v^ Rhbdf2^-/-^* mice (right). Notably, *Rhbdf1^v/v^ Rhbdf2^-/-^* mice develop a wavy hair coat (arrow). b. H&E-stained sections of skin from adult *Rhbdf1^v/+^ Rhbdf2^-/-^* and *Rhbdf1^v/v^ Rhbdf2^-/-^* mice. The *Rhbdf1^v/v^ Rhbdf2^-/-^* skin displays follicular dystrophy (arrow), hyperplasia (*), and hyperkeratosis (arrowhead), whereas that of *Rhbdf1^v/+^ Rhbdf2^-/-^* mice does not display these phenotypes. Top panels: low magnification; bottom panels: high magnification. c. *Rhbdf1^v/v^ Rhbdf2^-/-^* double mutant mice displayed an "eyelids open at birth" phenotype in 3-day-old mouse pups (arrow). No such defects were observed in *Rhbdf1^v/+^ Rhbdf2^-/-^* mice. Top panels: low magnification; bottom panels: high magnification. d. No cardiac or brain abnormalities were observed in either *Rhbdf1^v/v^ Rhbdf2^-/-^* double mutant mice or *Rhbdf1^v/+^ Rhbdf2^-/-^* mice. **Figure S7.** Non-AUG protein coding transcripts of *Rhbdf1*: a. Mouse non-AUG protein-coding *Rhbdf1* transcript (Top). Human non-AUG protein-coding *RHBDF1* transcript. b. C-terminal Myc-DDK-tagged empty vector (lanes 1, 2), *non-AUG* vector (lanes 3, 4) or a positive control vector (798-bp; lanes 5, 6) were transiently expressed in 293T cells, and cell lysates were analyzed using western blotting with FLAG-specific antibody. Blots were washed, blocked in 5% nonfat dry milk, and re-probed with anti-actin antibody. **Figure S8.** Retention of *En2* splice acceptor site in *Rhbdf1^v2/v2^* allele:a. mRNA expression of the *Rhbdf1* gene assessed by RT-qPCR using SYBR Green RT-PCR master mix and, exon 16 forward and exon 17 reverse primers. There is no difference in the mRNA levels of *Rhbdf1* between *Rhbdf1^+/+^* and *Rhbdf1^v2/v2^* mice. b. Schematic representation of *Rhbdf1^v2/v2^* allele indicating the presence of *En2* splice acceptor site (Extra) between exons 3 and 12. RT-PCR of spleens from *Rhbdf1^v2/v2^* mutant mice using primers in exon 3 and exon 12. Exon 3 forward (2817 For) and exon 12 reverse (2818 Rev) primers generated a 324-bp product, instead of an expected 209-bp product due to splicing of exons 3 and 12 together. c. Sanger's sequencing indicated the presence of *En2* splice acceptor site (Extra) flanked by exons 3 and 12. d. Schematic representation of the targeted KO-first allele used to generate ES cell clones displaying the splice acceptor site downstream of exon 3. e. Transcription and translation from the conventional start site in exon 2 can yield a 10.8 kDa truncated protein owing to a stop codon in the splice acceptor site. **Figure S9.** 5' RACE identifies alternative exon upstream of the conventional exon 1: a. Schematic representation of the *Rhbdf1* KO-first allele used to generate ES cell clones displaying an alternative exon 1 (arrow) upstream of the conventional exon 1. b. Schematic representation of the *Rhbdf1* genomic DNA displaying the alternative exon 1 (arrow) identified by 5' RACE. The DNA sequence of the alternative exon 1 is also shown. **Figure S10.** Translation initiation from alternative exon 1 in the *Rhbdf1^v2/v2^* allele could generate a truncated protein: a. Schematic representation of an *Rhbdf1^v2/v2^* splice variant displaying transcription initiation from the alternative exon 1, which could result in a 285-bp transcript and a 12.3 kDa truncated protein. b. Primers (2827, 2818, and 2826) used to sequence verify splicing of alternative exon 1 with exon 2 (top), and *En2* splice acceptor site (*En2* SA) with exon 12 (bottom). **Figure S11.** Splicing of alternative exon 1 with exon 4 generates an N-terminally truncated variant 1 protein: a. Schematic representation of an *Rhbdf1^v2/v2^* splice variant displaying transcription initiation from the exon 14, which could result in a 801-bp transcript and a 32.77 kDa N-terminally truncated protein (variant 1). b. RT-PCR of spleens from *Rhbdf1^v2/v2^* mutant mice using primers (2827 and 2814) in alternative exon 1 and exon 15 resulted in a 430-bp amplicon. Sanger's sequence demonstrated splicing of alternative exon 1 with *En2* SA. Notably, *Rhbdf1^v/v^* mutant mice resulted in an amplicon with primers 2827 and 2814, suggesting the use of alternative exon 1. However, higher molecular weight PCR fragments in *Rhbdf1^v/v^* mutant mice can be attributed to the presence of exons 4 through 11, which are excised in *Rhbdf1^v2/v2^* mutant mice. Thermocycling conditions using *Taq* DNA Polymerase with standard buffer are also shown. **Figure S12.** Splicing of alternative exon 1 with exon 4 can also generate an N-terminally truncated variant 2 protein: a. Schematic representation of an *Rhbdf1^v2/v2^* splice variant displaying transcription initiation from the exon 15, which could result in a 717-bp splice variant and a 29.6 kDa N-terminally truncated protein (variant 2). b. RT-PCR of spleens from *Rhbdf1^v2/v2^* mutant mice using primers (2827 and 2814) in alternative exon 1 and exon 15 also resulted in a 639-bp amplicon. Sanger's sequence demonstrated splicing of *En2* SA with exon 12. **Figure S13.** Double digests with NEB's restriction enzymes: Double digests of C-terminal Myc-DDK-tagged vectors (1 μg DNA) with NEB's restriction enzymes EcoRI-HF and NotI-HF. **Figure S14.** Unpredicted reinitiation of transcription and translation: Our data strongly suggests that mice harboring

Hosur et al. Genome Biology   (2020) 21:168

Page 23 of 25

homozygous definitive-null or homozygous targeted KO-first alleles are homozygous null, however, mice harboring homozygous KO-first mutant alleles *tm1b* and *tm1d* can be null or gain-of-function or hypomorphic mice (arrows) owing to unpredicted reinitiation of transcription and translation. **Figure S15.** PCR and genotyping of *Rhbdf1* and *Rhbdf2* KOMP alleles: a. PCR primers used to genotype *Rhbdf1*$^{-/-}$ mice (A5) generated in the present study. Thermocycling conditions using *Taq* DNA Polymerase with standard buffer. b. Schematic showing the PCR strategy and the primers used to genotype *Rhbdf2*$^{-/-}$ mice. Thermocycling conditions using *Taq* DNA Polymerase with standard buffer. PCR products were purified and sequence verified.

Additional file 2. Table S1. Incomplete CDS mouse transcripts.

Additional file 3. Table S2. Incomplete CDS mouse genes.

Additional file 4. Table S3. Incomplete CDS human transcripts.

Additional file 5. Table S4. Incomplete CDS human genes.

Additional file 6. Table S5. Genes retaining residual protein expression upon CRISPR/Cas9-induced indels.

Additional file 7. Table S6. Bioinformatic analyses of conditional-ready, lethal, and subviable strains.

Additional file 8. Table S7. Bioinformatic analyses of targeted KO-first strains exhibiting conflicting phenotypes.

Additional file 9. Review history.

Authors' contributions

VH, LDS, and MVW designed and supervised the project. VH, BEL, and DL performed the experiments. GAC and VK contributed to bioinformatic analyses. VH, LDS, and MVW analyzed the data and wrote the manuscript. All authors critically reviewed and approved the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. Exome-seq and RNA-seq data are available under the NCBI BioProject: PRJNA641484 [41].

Ethics approval and consent to participate

The Animal Care and Use Committee (Animal Use Summary # 11006) at The Jackson Laboratory approved all experimental procedures.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]The Jackson Laboratory for Mammalian Genetics, Bar Harbor, ME, USA. [2]Arthritis and Tissue Degeneration Program, Hospital for Special Surgery at Weill Cornell Medicine, New York, NY 10021, USA.

References

1. Christova Y, Adrain C, Bambrough P, Ibrahim A, Freeman M. Mammalian iRhoms have distinct physiological functions including an essential role in TACE regulation. EMBO Rep. 2013;14:884–90.

2.   Issuree PD, Maretzky T, McIlwain DR, Monette S, Qing X, Lang PA, Swendeman SL, Park-Min KH, Binder N, Kalliolias GD, et al. iRHOM2 is a critical pathogenic mediator of inflammatory arthritis. J Clin Invest. 2013;123:928–32.
3.   Yan Z, Zou H, Tian F, Grandis JR, Mixson AJ, Lu PY, Li LY. Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. Mol Cancer Ther. 2008;7:1355–64.
4.   Zhou Z, Liu F, Zhang ZS, Shu F, Zheng Y, Fu L, Li LY. Human rhomboid family-1 suppresses oxygen-independent degradation of hypoxia-inducible factor-1alpha in breast cancer. Cancer Res. 2014;74:2719–30.
5.   Xu Q, Chen C, Liu B, Lin Y, Zheng P, Zhou D, Xie Y, Lin Y, Guo C, Liu J, Li L. Association of iRhom1 and iRhom2 expression with prognosis in patients with cervical cancer and possible signaling pathways. Oncol Rep. 2020;43:41–54.
6.   Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, et al. The knockout mouse project. Nat Genet. 2004;36:921–4.
7.   Brown SD, Moore MW. Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. Dis Model Mech. 2012;5:289–92.
8.   Collins FS, Finnell RH, Rossant J, Wurst W. A new partner for the international knockout mouse consortium. Cell. 2007; 129:235.
9.   Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling MC, Bottomley J, Brown SD, Burger A, Bult CJ, et al. the mammalian gene function resource: the International Knockout Mouse Consortium. Mamm Genome. 2012;23:580–6.
10. Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 2011;474:337–42.
11. White JK, Gerdin AK, Karp NA, Ryder E, Buljan M, Bussell JN, Salisbury J, Clare S, Ingham NJ, Podrini C, et al. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell. 2013;154:452–64.
12. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013;153:910–8.
13. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816–21.
14. Makino S, Fukumura R, Gondo Y. Illegitimate translation causes unexpected gene expression from on-target out-of-frame alleles created by CRISPR-Cas9. Sci Rep. 2016;6:39608.
15. Smits AH, Ziebell F, Joberty G, Zinn N, Mueller WF, Clauder-Munster S, Eberhard D, Falth Savitski M, Grandi P, Jakob P, et al. Biological plasticity rescues target activity in CRISPR knock outs. Nat Methods. 2019;16:1087–93.
16. Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J, Mathews DH, Kim J, et al. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. Nat Commun. 2019;10:4056.
17. Anderson JL, Mulligan TS, Shen MC, Wang H, Scahill CM, Tan FJ, Du SJ, Busch-Nentwich EM, Farber SA. mRNA processing in mutant zebrafish lines generated by chemical and CRISPR-mediated mutagenesis produces unexpected transcripts that escape nonsense-mediated decay. PLoS Genet. 2017;13:e1007105.
18. Li X, Maretzky T, Weskamp G, Monette S, Qing X, Issuree PD, Crawford HC, McIlwain DR, Mak TW, Salmon JE, Blobel CP. iRhoms 1 and 2 are essential upstream regulators of ADAM17-dependent EGFR signaling. Proc Natl Acad Sci U S A. 2015;112:6080–5.
19. Adrain C, Zettl M, Christova Y, Taylor N, Freeman M. Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. Science. 2012;335:225–8.
20. Hosur V, Johnson KR, Burzenski LM, Stearns TM, Maser RS, Shultz LD. Rhbdf2 mutations increase its protein stability and drive EGFR hyperactivation through enhanced secretion of amphiregulin. Proc Natl Acad Sci U S A. 2014;111:E2200–9.
21. McIlwain DR, Lang PA, Maretzky T, Hamada K, Ohishi K, Maney SK, Berger T, Murthy A, Duncan G, Xu HC, et al. iRhom2 regulation of TACE controls TNF-mediated protection against Listeria and responses to LPS. Science. 2012;335:229–32.
22. Du X, Tabeta K, Hoebe K, Liu H, Mann N, Mudd S, Crozat K, Sovath S, Gong X, Beutler B. Velvet, a dominant Egfr mutation that causes wavy hair and defective eyelid development in mice. Genetics. 2004;166:331–40.
23. Blobel CP. ADAMs: key components in EGFR signalling and development. Nat Rev Mol Cell Biol. 2005;6:32–43.
24. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, Bennett R, Bhai J, Billis K, Boddu S, et al. Ensembl 2019. Nucleic Acids Res. 2019;47:D745–d751.
25. Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol. 1989;9:5073–80.
26. Gordon K, Futterer J, Hohn T. Efficient initiation of translation at non-AUG triplets in plant cells. Plant J. 1992;2:809–13.
27. Peabody DS. Translation initiation at non-AUG triplets in mammalian cells. J Biol Chem. 1989;264:5031–5.
28. Diaz de Arce AJ, Noderer WL, Wang CL. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. Nucleic Acids Res. 2018;46:985–94.
29. Lalonde S, Stone OA, Lessard S, Lavertu A, Desjardins J, Beaudoin M, Rivas M, Stainier DYR, Lettre G. Frameshift indels introduced by genome editing can lead to in-frame exon skipping. PLoS One. 2017;12:e0178700.
30. Mou H, Smith JL, Peng L, Yin H, Moore J, Zhang XO, Song CQ, Sheel A, Wu Q, Ozata DM, et al. CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. 2017;18:108.
31. Marcel Kapahnke, Antje Banning, Ritva Tikkanen. Random Splicing of Several Exons Caused by a Single Base Change in the Target Exon of CRISPR/Cas9 Mediated Gene Knockout. Cells. 2016;5(4):45.
32. Birling M-C, Yoshiki A, Adams DJ, Ayabe S, Beaudet AL, Bottomley J, Bradley A, Brown SD, Bürger A, Bushell W, et al. A resource of targeted mutant mouse lines for 5,061 genes. bioRxiv. 2019:844092.
33. Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H, et al. High-throughput discovery of novel developmental phenotypes. Nature. 2016;537:508–14.
34. Kochetov AV. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. Bioessays. 2008;30:683–91.
35. Huin V, Buee L, Behal H, Labreuche J, Sablonniere B, Dhaenens CM. Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer's disease and progressive supranuclear palsy brains. Sci Rep. 2017;7:12589.
36. Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. 2017;31:1717–31.
37. Dolgin E. Mouse library set to be knockout. Nature. 2011;474:262–3.
38. Gurumurthy CB, O'Brien AR, Quadros RM, Adams J Jr, Alcaide P, Ayabe S, Ballard J, Batra SK, Beauchamp MC, Becker KA, et al. Reproducibility of CRISPR-Cas9 methods for generation of conditional mouse alleles: a multi-center evaluation. Genome Biol. 2019;20:171.

Hosur *et al. Genome Biology*   (2020) 21:168

Page 25 of 25

39. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32:279–84.

40. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013;155:1479–91.

41. Hosur V, Low BE, Li D, Stafford GA, Kohar V, Shultz LD, Wiles MV: Genes adapt to outsmart gene targeting strategies in mutant mouse strains by skipping exons to reinitiate transcription and translation. Dataset. National Center for Biotechnology Information: PRJNA641484 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.