Supporting Appendix

Supplementary Materials and Methods

Engineering of an optimal COIN module

An optimal COIN module was developed by testing different 3'SS and pA combinations using an artificial mini-gene that can express either dsRED2 or eGFP by alternative splicing. This mini-gene was engineered by modifying the single exon artificial gene encoding dsRED2 in plasmid pDsRed2-N1 (Clonetech). This choice was made because transient transfection of pDsRed2-N1 in mammalian cell lines supports a high level of expression of dsRED2, and therefore enables stringent testing of any read-through past potential COIN modules. To introduce and test different potential COIN modules, the dsRED2-coding exon was first split into two exons using the rabbit beta hemoglobin (HBB2) intron 2 (Supporting Appendix, Fig. S1). This resulted in a *dsRED2* comprised of two exons and in which the first exon is a shared exon composed of mostly 5'UTR sequence and encoding only the first three amino acids of dsRED2 (Met-Ala-Ser; with the corresponding DNA sequence in the exon being ATGGCCAG). In this manner, exon 1 could also function as a shared initial exon for eGFP. Splitting dsRED2 into two exons had no apparent effect on its expression (Supporting Appendix, Fig. S1*B*, panel iii).

To test the ability of different potential COIN modules to function as efficient terminal exons, the split dsRED2 mini-gene was further modified by introducing a *[lox66]-3'SS-eGFP-lox71* (*3'SS-eGFP*) element at the *MfeI* site of the intron (Supporting Appendix, Fig. S1*A*). The *eGFP* portion of *3'SS-eGFP* initiates its open reading frame by using exon 1 of *dsRED2* (and is in the same phase — phase 1 — as the shared exon). The 3'SS chosen was also derived from HBB2 intron 2, as described above (the specific sequence being 5'-CGG GCC CCT CTG CTA ACC ATG TTC ATG CCT TCT TCT TTT TCC TAC AG-3'). Inclusion of the *3'SS-eGFP* element alone, devoid of a polyadenylation region (pA) after eGFP's stop codon, was not adequate for the *3'SS-eGFP* element to function as a terminal exon, as evidenced by co-expression of both dsRED2 and eGFP from the corresponding plasmid

(pR2i[(L66)/rßgl3'SS-GFP/L71]; see Figure S1*B*, panel iv); this observation is consistent with the coupling that occurs between splicing and polyadenylation (1). Two different pA regions, bovine growth hormone (bGH) pA (2) and HBB2 pA (3), were then tested their ability to convert the *3'SRSS-eGFP* element into a terminal exon, both in the context of the rabbit beta-globin 3'SS and the Adenovirus major late transcript 3' splice site (Adml3'SS) (2, 4, 5). This comparison revealed that the rabbit beta-globin pA was more efficient in converting the *3'SS-eGFP* element into a terminal exon, as evidenced by the absence of any dsRED2-expressing cells in cells transfected with plasmids pR2i[(L66)/rßgl3'SS-GFP-ßglpA/L71] and pR2i[(L66)/Adml3'SS-GFP-ßglpA-L71] (Supporting Appendix, Fig. S1*B*, panels vi and viii). In contrast, when the bGH pA was utilized, a significant number of cells expressed both dsRED2 and eGFP (Supporting Appendix, Fig. S1*B*, panels v and vii). Lastly, the 3'SS of *En2* was tested in combination with SV40pA (6). Although this pair was very efficient at terminating transcription after *eGFP*, altogether abrogating *dsRED2* expression, it was deemed unsuitable for the COIN module for two reasons: it reduced the level of eGFP expression (Supporting Appendix, Fig. S1*B*, panels ix and x); and, the SV40pA region contains polyadenylation signals on both strands (7). Therefore, the rßgl3'SS-GFP-ßglpA element was chosen for the COIN module, because it functioned as an efficient terminal exon in this assay. Although the Adml3'SS-GFP-ßglpA element worked equally well, the choice was based on the idea of retaining a 3'SS-pA pair that has co-evolved.

ES cell targeting

In preparation for ES cell targeting, purified BACvec targeting vector DNA was analyzed for integrity by Pulse Field Gel Electrophoresis as described (8). 1.5 µg of targeting vector was dissolved in 50 µl of EmbryoMax® electroporation buffer (EMD Millipore, catalog number ES-003-D) and mixed with 75 µl of a cell suspension containing 7.5 million of VGF1 ES cells (9) (129S6SvEvTac/C57BL6NTac) in electroporation buffer. The mixture was subjected to electroporation as described (10), except that a multi-well electroporation device (Harvard Apparatus, Boston,

MA) was utilized, followed by a 10 minute incubation on ice before transfer to a 15 cm gelatinized plate. After 48 hours of growth at 37°C in a CO_2 incubator, drug-resistant colonies were selected by growth for an additional eight days in medium containing G418 (0.1 mg/ml) or Hygromycin B (0.2 mg/ml), with medium changed daily. Colonies were picked and transferred to individual wells of a 96-well plate containing trypsin, followed by transfer to a gelatinized 96-well plate. After three days of growth, the cells were disrupted with trypsin and half of each well was frozen while the other half was transferred to a new plate for genotyping.

Hprt1^{ex3COIN} allele

The ex3COIN allele of *Hprt1* (VG1272) was engineered by introducing the COIN intron into exon 3 of *Hprt1* (Ensembl exon ID: ENSMUSE00000491684) (Fig. 2*A*) after the 85th nucleotide of that exon (Supporting Appendix, Fig. S2), corresponding to nucleotide 50,355,349 of Chromosome X, hence splitting exon 3 into two exons, 3L and 3R, 85 and 99 bp respectively (Table 1). The targeting vector was generated using BAC RP23-13n1, which contains ~201 kb of mouse genomic DNA encompassing *Hprt1*, by inserting the COIN intron (eGFP, phase 0; *neo*) into exon 3, and resulting in a BACVec with homology arms of ~116 and 85 kb. For targeting, this BACVec was linearized with *NotI*, and electroporated into F1H4 ES cells to generate allele VG1186 (*Hprt1ex3*$^{COIN(neo)}$). 9 out of 288 ES clones were correctly targeted. Germline-transmitting clones 1186D-A12 and 1186D-G2 were prepared for experiments in ES cells by excision of *neo* using FLPe to generate *Hprt1*^{ex3COIN}*/Y* (VG1272) ES cells, and subsequent treatment with Cre to generate *Hprt1*$^{ex3COIN\text{-}INV}$*/Y* ES cells. A detailed protocol for 6-TG treatment of *Hprt1* ES cell lines is provided in the Supplementary Appendix.

Il2Rg^{ex1COIN} alleles

Two ex1COIN alleles of *Il2Rg* were generated: an ex1COIN allele (*Il2rg*^{ex1COIN}) and a 'reverse' ex1COIN allele (*Il2rg*^{e1REVCOIN}); the latter is identical to the former except that the COIN module is placed in the sense strand, thereby providing a null

allele of *Il2rg* that can be converted back to WT. The ex1COIN allele of *Il2rg* (VG1253) was generated by inserting the COIN intron into exon 1 of *Il2rg* (Ensembl exon ID: ENSMUSE00000806649) (Fig. 3*A*) after the 89th nucleotide of that exon (Supporting Appendix, Fig. S3), corresponding to nucleotide # 98,463,506 of Chromosome X, splitting exon 3 into exons, 3L and 3R, 89 and 112 bp respectively (Table 1). To generate the targeting vector, BAC 290o15 containing ~97kb of mouse genomic DNA encompassing *Il2rg* was selected from a BAC library of 129/Svj mouse genomic DNA (Incyte – Release II). BAC 290o15 was modified by insering the COIN intron (eGFP, phase 0; *neo*) into exon 1 of *Il2rg*, resulting in a BACVec with homology arms of ~7 and 90 kb. For targeting, this BACVec was linearized with *NotI*, and electroporated into F1H4 ES cells. 11 out of 288 ES clones were correctly targeted. Clone 1202A-F11 ($Il2rg^{ex1COIN(neo)/+}$; VG1202) was treated with FLPe to remove the DSC and give rise to $Il2rg^{ex1COIN/+}$ ES cells (VG1253). Clones 1253A-B9 and 1253B-A5 were used to generate $Il2rg^{ex1COIN/+}$ mice using VelociMouse technology. The $Il2rg^{e1REVCOIN}$ allele was engineered in an identical manner except that COIN module was placed in the sense strand. The flow cytometry protocol utilized in phenotyping *Ilr2g* lines is provided in the Supplementary Appendix.

Dll4^{i3COIN} alleles

Two i3COIN alleles were engineered for *Dll4*, VG1407 and VG1513, differing only in that VG1407 incorporates TMeGFP as a reporter, whereas VG1513 incorporates TMT2AeGFP, both in phase 1. For both alleles, the COIN module was inserted into intron 3 of *Dll4*, coordinates 119,153,488 to 119,153,520 on chromosome 2, while simultaneously deleting a 33 bp non-conserved region (AGC AAG ATT CTG GGA CCC TCT GGC CTG TGC TTA) to accommodate LOA probes, and effectively divide the 808 bp intron 3 into a 5' and 3' part of 631 and 144 bp respectively (Table 1; Fig. 4*A*), while increasing the total size of intron 3 to 2,532 bp for VG1407 and 2,589 bp for VG1513. To engineer targeting vectors, BAC 475d4 containing ~138 kb of mouse genomic DNA encompassing *Dll4* was selected by PCR-screen from a BAC library of 129/Svj mouse genomic DNA (Incyte – Release II). To generate alleles VG1407 ($Dll4^{i3COIN}$) and VG1513 ($Dll4^{i3T2ACOIN}$), each COIN module

and accompanying DSC were introduced into BAC 475d4 in intron 3 of *Dll4* using BHR, resulting in BACVecs with ~115 and ~25 kb homology arms. For targeting, each BACVec was linearized with *NotI*.

To generate VG1407, the corresponding BACVec for was electroporated into F1H4 ES cells, to generate targeted clones for VG1361 (*Dll4*$^{i3COIN(neo)/+}$). 11 out of 192 ES clones were correctly targeted. ES cell clone 1361A-B5 was targeted at *Gt(ROSA)26Sor* to introduce CreER^{t2} (11), and then one of the resulting targeted clones, 2192N-B12 (*Dll4*$^{i3COIN(neo)/+}$; *Gt(ROSA)26Sor*$^{CreERt2\text{-}hyg/+}$), was treated with FLPe to remove the DSCs and generate clone 1407B-B4 (*Dll4*$^{i3COIN/+}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$), which was then used to generate *Dll4*$^{i3COIN/+}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$ mice using VelociMouse technology.

To generate VG1513, the corresponding BACVec for was electroporated into F1H4 2192L-F4 ES cells (which are *Gt(ROSA)26Sor*$^{CreERt2\text{-}hyg/+}$) to generate targeted clones for VG1495 (*Dll4*$^{i3T2ACOIN(neo)/+}$; *Gt(ROSA)26Sor*$^{CreERt2\text{-}hyg/+}$). ES cell clones 1495A-B7 and 1495A-H6 were subsequently treated with FLPe to remove the *neo* and *hyg* and give rise to VG1513 ES cells. Clone 1513B-H2 was used to generate *Dll4*$^{i3T2ACOIN/+}$; *Gt(ROSA)26Sor*$^{CreERt2\text{-}hyg/+}$ mice using VelociMouse technology. Protocols for genotyping *Dll4* lines as well as analysis of their retinas are provided in the Supplementary Appendix.

Drosha null allele (*Drosha*$^{LacZ\text{-}neo}$; VG549)

The ***Drosha***$^{LacZ\text{-}neo}$ allele was engineered by replacing coordinates 12,820,288 (end of exon 22 in Drosha splice isoform ENSMUST00000090292), to 12,860,423 (within intron 34 in ENSMUST00000090292) of Chromosome 15 with *LacZ-neo*.

Drosha^{ex4COIN} allele

The ex4COIN allele of *Drosha* (VG1483) was generated by inserting the COIN intron into exon 4 (splice isoform ENSMUST00000090292, Ensembl exon ID: ENSMUSE00000563117) (Fig. 5*A*) after the 625th nucleotide of that exon (Supporting Appendix, Fig. S5), corresponding to nucleotide 12,764,208 of

Chromosome 15, and splitting exon 4 into exons, 4L and 4R, 625 and 209 bp respectively (Table 1). The targeting vector was generated using BAC RP23-78l8, by inserting the COIN intron (eGFP, phase 0; *hyg*) into exon 4 (eGFP, phase 0; *hyg*), to generate a BACVec with homology arms of ~111 and 60 kb. In preparation of targeting, this BACVec was linearized with *PI-SceI*, and electroporated into 549C-C2 and 549C-H4 ES cells to generate targeted ES cells for allele VG1458 (*Drosha*$^{ex4COIN(hyg)}$). 26 out of 384 ES clones were correctly targeted; two of these, 1458A-E7 and 1458C-A1, were treated with FLPe to remove *hyg* and generate *Drosha*$^{ex4COIN/LacZ\text{-}neo}$ ES cells. *Drosha*$^{ex4COIN/LacZ\text{-}neo}$ clone 1483A-F10 was then retargeted with CreER^{t2} into *Gt(ROSA)26Sor* to generate *Drosha*$^{ex4COIN/LacZ\text{-}neo}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$ ES cells. The experiments described here were conducted using clone 2192AB-A6.

Drosha^{i3COIN} allele

The i3COIN allele of *Drosha* (VG1390) was engineered by introducing the COIN module into intron 3 of splice isoform ENSMUST00000090292) between coordinates 12,762,596 to 12,762,658 on chromosome 15 while simultaneously deleting the corresponding 63 bp non-conserved region (TCC CTG GCC TTC AGG GAA TGA AAT GTT GGT TCC TCC TAA GTC CGA GGC ATT GCC TGG TGC CCA) to accommodate LOA probes, dividing the 6,536 bp intron 3 into a 5' and 3' part of 5,548 and 925 bp respectively (Table 1), and increasing the total size of intron 3 to 8,112 bp. The targeting vector was generated using BAC RP23-78l8, which contains ~171 kb of mouse genomic DNA encompassing *Drosha*, by inserting the COIN module (eGFP, phase 2; *hyg*) into intron 3 resulting in a BACVec with homology arms of ~112 and 59 kb. For targeting, this BACVec was linearized with *PI-SceI*, and electroporated into 549C-C2 and 549C-H4 ES cells, to generate allele VG1350 (*Drosha*$^{i3COIN(hyg)}$). 31 out of 384 ES clones were correctly targeted. Clones 1350D-B1 and 1350D-C12 (*Drosha*$^{i3COIN(hyg)/LacZ\text{-}neo}$) were treated with FLPe to remove *hyg*, and generate allele VG1390 (*Drosha*^{i3COIN}). *Drosha*$^{i3COIN/LacZ\text{-}neo}$ clones 1390B-B1 and

1390D-D12 *were* retargeted to introduce CreER^{t2} into *Gt(ROSA)26Sor* to generate *Drosha*$^{i3COIN/LacZ\text{-}neo}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$ ES cells.

Chordin^{i2COIN} allele

An i2COIN allele of *Chordin* was generated by inserting the COIN module into intron 2 of *Chordin* between coordinates 20,733,914 to 20,733,974 on chromosome 16 while simultaneously deleting the corresponding 61 bp non-conserved region (ATA AAT GAC ATT ATT AGT TCT ACT TTG CAG ACA GGG ACG CCG AGG CTC AGA GAA GTT AAG A) to accommodate LOA probes, dividing the 372 bp intron 2 into a 5' and 3' part of 206 and 105 bp respectively (Table 1), and increasing the total size of intron 2 to 2,068 bp. The targeting vector was generated using BAC RP23-148p12, which contains ~211 kb of mouse genomic DNA encompassing *Chordin*, by inserting the COIN module (TMeGFP, phase 0; *hyg*) into intron 2, and resulting in a BACVec with ~83 and ~127 kb homology arms. For targeting, this BACVec was linearized with *PI-SceI*, and electroporated into F1H4 ES cells. 12 out of 288 ES clones were correctly targeted. Clones 1231C-E7 and 1231C-F2 were used to generate *Chordin*$^{i2COIN(hyg)/+}$ mice.

Ctgf^{ex2COIN} allele

The ex2COIN allele of *Ctgf* (VG1511) was generated by inserting the COIN intron into exon 2 of *Ctgf* at coordinate 24,315,801 of chromosome 10, effectively splitting the 223 bp exon into left (2L) and right (2R) exons of 120 and 103 bp respectively (Table 1). To generate the targeting vector, BAC 460d11 containing ~160 kb of mouse genomic DNA encompassing *Ctgf* was selected from a BAC library of 129/Svj mouse genomic DNA (Incyte – Release II). The COIN modification cassette (TMeGFP, phase 0; *hyg*) was introduced into BAC 460d11 in exon 2 of *Ctgf*, resulting in a BACVec with ~120 and ~40 kb homology arms. For targeting, this BACVec was linearized with *NotI*, and electroporated into F1H4 286A-B8 ES cells that already harbor a null allele of *Ctgf* (12). 12 out of 192 ES clones were correctly targeted. Clones 1367B-C8, 1367B-F10, 1367B-F12, and 1367B-G6 were used to generate *Ctfgf*$^{ex2COIN(hyg)/LacZ\text{-}neo}$ mice, which where then bred to a Flp deletor to

generate *Ctfgf*$^{ex2COIN/+}$. Phenotypic characterization of these mice has been published (13, 14).

Dicer1 alleles

A null allele for *Dicer1* was engineered by replacing nucleotides 105,946,730 to 105,986,312 on chromosome 12 (a 39.6 kb region encompassing the coding sequence contained in exons 3 to 27; see Dataset S1 and Supporting Appendix, Fig. S7) with LacZ. To generate the targeting vector, BAC 254h1 containing ~140 kb of mouse genomic DNA encompassing *Dicer1* was selected by PCR-screen from a BAC library of 129/Svj mouse genomic DNA (Incyte – Release II). The coding region of Dicer1 from amino acid 52 to its stop codon was replaced with *LacZ-neo*, in a manner such that the ORF of LacZ (minus the ATG) begins after amino acid 52 of Dicer1. Expression of neomycin phosphotransferase in the floxed DSC was driven by the UBC promoter region (Supporting Appendix, Fig. S16). For targeting, the modified BAC was linearized with *NotI* to generate a linear targeting vector with homology arms of ~75 and ~25 kb flanking the *LacZ-neo* cassette, and was electroporated into ES line F1H4. 2 out of 192 ES clones were correctly targeted. Clone 548A-A4 was microinjected to generate mice, and also utilized to generate the i6COIN allele of *Dicer1*.

The i6COIN allele of *Dicer1* (VG1399) was generated by inserting the COIN module into intron 6 of *Dicer1* between coordinates 105,953,342 to 105,953,387 on chromosome 12, simultaneously deleting a 46 bp non-conserved region (AAA GA GCC TTC GTC AAA GCG GAG AGG AGA CAT GCT GGA GTT CCC TG) to accommodate LOA probes, effectively dividing the 4,162 bp intron 6 into a 5' and 3' part of 3,770 and 406 bp respectively (Table 1 and Supporting Appendix, Fig. S7), and increasing the size of intron 6 to 5,812 bp. The COIN intron (*eGFP*; phase 2; *hyg*) was introduced into BAC 254h1 in intron 6 of *Dicer* using BHR, resulting in a BACVec with ~83 and ~57 kb homology arms. For targeting, this BACVec was linearized with *NotI*, and electroporated into F1H4 548A-A4 ES cells. 12 out of 192 ES clones were correctly targeted. Clones 1353A-A2 and 1353A-D4 with genotype *Dicer1*$^{i6COIN(hyg)/LacZ\text{-}neo}$ were treated with FLPe to generate *Dicer1*$^{i6COIN/LacZ\text{-}neo}$ ES cells,

which were subsequently microinjected to generate mice using VelociMouse technology (9).

Gdf11^{ex1COIN} allele

The ex1COIN allele of *Gdf11* (VG1422) was engineered by inserting the COIN intron into exon 1 of *Gdf11*, at coordinate 128,328,478 in the antisense strand of chromosome 10, splitting the 472 bp exon into left (1L) and right (1R) exons of 297 and 175 bp respectively (Dataset S1 and Supporting Appendix, Fig. S8). The targeting vector was generated using BAC RP23-476c23 containing ~210 kb of mouse genomic DNA encompassing *Gdf11* inserting the COIN intron (TMeGFP, phase 0; *hyg*) into exon 1 of *Gdf11* at the position listed above, and resulting in a BACVec with homology arms of ~94.5 and ~39 kb. For targeting, this BACVec was restricted with *SalI*, and was electroporated into F1H4 ES cells to generate allele VG1381 (*Gdf11*$^{ex1COIN(hyg)}$). 15 out of 192 ES clones were correctly targeted. Clones 1381A-A3 and 1381A-A6 (*Gdf11*$^{ex1COIN(hyg)/+}$) were treated with FLPe to remove the FRTed *hyg* and give rise to VG1422 ES cells (*Gdf11*$^{ex1COIN/+}$). Clone 1422B-B1 was used to generate *Gdf11*$^{ex1COIN/+}$ mice using VelociMouse technology.

Gpr124 alleles

The knockout allele employed in our studies has been described elsewhere (15). The ex1COIN allele of *Gpr124* (VG1402) was engineered by inserting the COIN intron into exon 1 of *Gpr124* close to the start of the ORF, at coordinate 28196445 in the antisense strand of chromosome 8, effectively splitting the 371 bp exon into left (1L) and right (1R) exons of 133 and 238 bp respectively (Dataset S1 and Supporting Appendix, Fig. S9). The point of insertion was optimized to conform to the consensus sequence of the 5' splice site (16) by introducing a silent C to A silent mutation. To generate the targeting vector, a BAC containing mouse genomic DNA encompassing *Gpr124* was selected by PCR-screen from a BAC library of 129/Svj mouse genomic DNA (Incyte). The BAC id is 151h20 and it contains ~212 kb of genomic DNA. BAC 151h20 was modified by inserting the COIN intron (*eGFP*; phase 1; *hyg*) at the position listed above, resulting in a BACVec with homology arms of

~120 and 70 kb. For targeting, this BACVec was restricted with *NotI* and was electroporated into the F1H4-derived ES line 2251A-C6 (*Gt(ROSA26)Sor*$^{CreERt2/+}$), to generate allele VG804 (*Gpr124*$^{ex1COIN(hyg)/+}$). 13 out of 192 clones were targeted. One of these clones, 804A-D5 was treated with FLPe to remove the DSC and generate the final COIN allele, VG1402. Mice with were generated from clone 1402A-G12 (*Gpr124*$^{ex1COIN/+}$; *Gt(ROSA26)Sor*$^{CreERt2/+}$) and utilized in experiments.

Gt(ROSA26)Sor$^{i1COIN(dsRED2>eGFP)}$ allele

An i1COIN allele of *Gt(ROSA26)Sor* (VG2154) was generated by inserting the artificial *dsRED2 COIN* transgene contained in plasmid pR2i[(L66)/rßgl3'SR-GFP-ßglpA/L71] (Supporting Appendix, Fig. S1 and Supporting Appendix, Table S11) into the *XbaI* site of the intron 1 of *Gt(ROSA26)Sor*, by modifying a plasmid carrying 15.3 kb encompassing the *Gt(ROSA26)Sor* locus (Supporting Appendix, Fig. S10). At the insertion site a 2 bp deletion was simultaneously generated (removing nucleotides 113025822 to 113025824 on Chromosome 6) to accommodate LOA genotyping. For targeting, the targeting vector was linearized with *FseI*, resulting in a targeting fragment with homology arms of ~3.8 and ~11.5 kb. 20 out of 192 clones were targeted. ES cells clone 2154A-D1 was microinjected to generate mice, which were then bred to a global Cre deleter (17) to generate corresponding COIN-INV allele of VG2154 (*Gt(ROSA26)Sor*$^{i1COIN\text{-}INV(eGFP)}$).

Plxnd1 alleles

A knockout allele of *Plxnd1* was generated by replacing exons 2 to 8 with *TMLacZ-neo* (corresponding to nucleotides 115923787 to 115928751 of Chromosome 6). The ex2COIN allele of *Plxnd1* (VG1618) was engineered by inserting the COIN intron into exon 2 of *Plxnd1* (Ensembl exon ID: ENSMUSE00000306769) at coordinate 115928676 in the antisense strand of Chromosome 6, splitting the 177 bp exon into left (2L) and right (2R) exons of 97 and 80 bp respectively (Dataset S1 and Supporting Appendix, Fig. S11). The choice of introducing the COIN intron into exon 2 was in order to avoid disruption of a CpG island that straddles exon 1 (coordinates 115943412 to 115945151 on Chromosome 6). The targeting vector was generated

using BAC bMQ400h23, which contains ~165 kb of mouse genomic DNA encompassing *Plxnd1*, by inserting the COIN intron (*TM-T2A-eGFP*; phase 1; *hyg*) at the coordinates listed above, and resulting in a BACVec with homology arms of ~82 and 83 kb. For targeting, this BACVec was restricted with *SalI*, and was electroporated into the F1H4 ES cells to generate allele VG853 ($Plxnd1^{ex2COIN(hyg)}$). 16 out of 192 clones were targeted. One of these clones, 853A-C4 was microinjected to generate mice, which were subsequently bred to a FLPe deleter line (VG2220) to generate line VG1618 ($Plxnd1^{ex2COIN/+}$).

Sost alleles

A knockout allele of *Sost* was generated via KOMP (line VG10069). Briefly, the entire protein-coding sequence of *Sost* was replaced with *LacZ-neo* in a manner such that the initiating ATG of Sost became the initiating Methionine of LacZ. All the details regarding the generation of this allele can be found at http://www.velocigene.com/komp/detail/10069. $Sost^{LacZ-neo/+}$ mice were generated using VelociMouse technology.

The iCOIN allele of *Sost* (VG1445) was generated by inserting the COIN module into the intron of *Sost* between coordinates 101,828,087 to 101,828,135 on chromosome 11, simultaneously deleting a 49 bp non-conserved region (CTG CAT TGC ATG CTC CTT TGG GGT GGC CCT TTC CTC CAC ATC CCT TAC T) to accommodate LOA probes, thereby dividing the 2,492 bp intron into a 5' and 3' part of 938 and 1,504 bp respectively (Table 1 and Supporting Appendix, Fig. S12), and increasing the size of the intron to 4,199 bp. The targeting vector was generated using BAC RP23-252b10, which contains ~287 kb of mouse genomic DNA encompassing *Sost*, by inserting the COIN module (*eGFP*, phase 1; *neo*) into Sost's single intron at the coordinates listed above. The BACVec was prepared for targeting by restricting it with *FseI* plus *SgrAI*, resulting in operational homology arms of 36.2 and 68.7 kb, and was electroporated into F1H4 2251A-C6 ES cells ($Gt(ROSA)26Sor^{CreERt2}$) to generate allele VG1426 ($Sost^{iCOIN(neo)}$). 5 out of 192 ES clones were targeted. Clones 1426A-H11 and 1426B-E10 ($Sost^{iCOIN(neo)/+}$; $Gt(ROSA)26Sor^{CreERt2-hyg/+}$) were treated with FLPe to generate $Sost^{iCOIN/+}$;

Gt(ROSA)26Sor$^{CreERt2/+}$ ES cells. Clone 1445B-E4 was subsequently microinjected to generate mice using VelociMouse technology.

Tie1 alleles

A knockout allele of *Tie1* (VG1222) was engineered so that it mirrors a previously published allele (18). Briefly, the protein-coding sequence of exon 1 plus the first four nucleotides of *Tie1*'s intron 1 (ATG GTC TGG TGG GGA TCC TCT TTG CTG CTC CCC ACT CTT TTC TTG GCC TCT CAT GTT ggta, where intronic sequence is denoted by lower case) were replaced with *LacZ-neo*. To generate the targeting vector, BAC 342e12 containing ~120 kb of mouse genomic DNA encompassing *Tie1* was selected by PCR-screen from a BAC library of 129/Svj mouse genomic DNA (Incyte), and modified as described above. For targeting, the resulting BACVec was restricted with *NotI* to generate a linear targeting vector with BAC homology arms of approximately 80 and 60 kb flanking the *LacZ-neo* cassette, and was electroporated into ES line F1H4. 5 out of 288 were targeted. Clones 1222A-E4 and 1222B-E4 were microinjected to generate *Tie1*$^{LacZ\text{-}neo/+}$ mice.

The ex1COIN allele of *Tie1* (VG1260) was engineered by inserting the COIN intron into exon 1 of *Tie1* (ENSMUSE00000756638), at coordinate 118,162,287 of chromosome 4, splitting the 424 bp exon into left (1L) and right (1R) exons of 380 and 34 bp respectively (Dataset S1 and Supporting Appendix, Fig. S13). The targeting vector was engineered using BAC 342e12 inserting the COIN intron (*TMeGFP*, phase 0; *hyg*) in the coordinate listed above. For targeting, the resulting BACVec was restricted with *NotI*, and electroporated into F1H4 ES cells, to generate allele VG1197 (*Tie1*$^{ex1COIN(hyg)}$). 12 out of 288 ES clones were correctly targeted. Clones 1197A-G1 and 1197B-A2 were treated with FLPe to excise *hyg* and give rise to VG1260 ES cells (*Tie1*$^{ex1COIN/+}$). Clone 1260B-A11 was used to generate *Tie1*$^{ex1COIN/+}$ mice.

Tg(Nanog-Cre) Cre deleter line

The Nanog-Cre transgenic mouse line was derived by electroporation of a modified BAC carrying the mouse *Nanog* gene into F1H4 ES cells. The complete coding sequence for *Nanog* was deleted from the second in-frame ATG codon (chromosome 6, genome coordinate 122,657,876) and replaced with a sequence encoding the phage P1 Cre recombinase preceded by a sequence encoding an amino-terminal nuclear localization sequence. The Cre coding sequence was followed by the polyadenylation signal from the SV40 virus. The insertion also included a selection cassette, flanked by *FRT* sites, consisting of the promoter from the mouse *Pgk1* gene linked to the EM7 bacterial promoter followed by the coding sequence for neomycin phosphotransferase and the polyadenylation signal from the *Pgk1* gene. An ES cell clone containing a single-copy random integration of the Nanog-Cre modified BAC was microinjected into a Swiss Webster 8-cell stage embryo and transferred into a surrogate mother. A completely ES-cell-derived F0 generation founder was backcrossed to C57BL/6NTac to establish the transgenic Nanog-Cre line (VG2258).

Gt(ROSA26)Sor^{FLPe} FLPe deleter line

The *Gt(ROSA26)Sor^{FLPe}* FLPe deleter line (VG2220) was generated by inserting a CDNA encoding FLPe into the XbaI site of the intron 1 of *Gt(ROSA26)Sor*. DSC and LOA probes are as described above for *Gt(ROSA26)Sor$^{i1COIN(dsRED2>eGFP)}$*. The targeting vector was eletroporated into F1H4 ES cells and 14 out of 192 clones were targeted, indicating a targeting frequency of 7.3%. Clone 2220B-C8 was used to generate mice and establish the line.

6-Thioguanine-induced killing in *Hprt1^{ex3COIN}/Y* ES cells

Hprt1^{+}/Y (F1H4), *Hprt1^{ex3COIN}/Y*, and *Hprt1$^{ex3COIN-INV}$/Y* ES cells were seeded at 2 x 10^4 cells/10cm dish and cultured overnight under standard ES cell culture conditions. The following day the medium was changed to medium with and without 10 μM 6-Thioguanine (6-TG, Sigma). The 6-TG-containing medium was changed daily for 10 days. The cells were rinsed with PBS and fixed with 100% methanol (JT Baker), then stained with Giemsa solution (Harleco) diluted 1:50 into

50% methanol, rinsed with water to remove the excess stain, and allowed to dry overnight. The results were documented photographically.

Preparation of cells for flow cytometry

Single cell suspensions from bone marrow and spleen were collected using standard methods. For the preparation of cells for flow cytometry, nonspecific binding sites were blocked with anti-mouse CD16/32 (BD Pharmingen) prior to staining of the cells with specific combination of antibodies. The following antibody combinations (all from Pharmingen) were used for staining: For detection of IgM+, B220+ cells, rat anti-Mouse IgM-PE plus rat anti-mouse CD45R (B220); for detection of natural killer cells, rat anti-mouse CD49b (DX5)-PE plus rat anti-mouse CD11b-APC. All monoclonal antibodies were prepared in a mass dilution/cocktail and were added to a final concentration of $0.5mg/10^5$ cells. Cells were fixed with BD Cytofix (BD Pharmingen), according to manufacture's instructions, and final pellets were re-suspended in 0.5ml staining buffer and analyzed using BD FACS Calibur and BD CellQuest Pro software.

Retina analysis

Mouse pups were treated with tamoxifen at 25 mg/kg for 48 hours using single subcutaneous injection at P5. Retinas were harvested at P7 and processed essentially as described (19). Briefly, mouse pups were humanely euthanized, eyes were enucleated, retinas were dissected, fixed for 1 hour with 4% paraformaldehyde, stained with FITC-labeled *Griffonia simplicifolia* (GS) lectin I (Vector Laboratories, Burlingame, CA), and flat-mounted. For eGFP expression studies, animals were injected with tamoxifen at P2 and tissues were harvested at P6. Retinas were fixed for 15 min and skin samples taken from the belly were kept unfixed.

PCR-based COIN inversion genotyping assay

Inversion of the COIN module in exonic COINs was performed using a PCR-based genotyping assay. Reactions of 25µl total volume were set up, using ex-Taq

DNA polymerase (TaKaRa) in the buffer and nucleotide concentrations recommended by its manufacturer, and 0.5 μg of genomic DNA as the template. The non-inverted COIN allele was detected by scoring for the presence of a 338 bp band that is amplifiable from non-inverted exonic COINs (but not from inverted exonic COINs), using primers INVdiaF2 (CAC TTT CTA CTC TGT TGA C) and INVdiaR2 (CCT TAC ATG TTT TAC TAG). The cycling conditions for this reaction were an initial melting step (95°C, 1 min), followed by 35 cycles of 95°C for 1 min, 55°C for 45 secs, and 72°C for 45 secs. The inverted COIN allele was detected by scoring for the presence of a 600 bp band that is amplifiable from inverted exonic COINs (but not from non-inverted exonic COINs), using primers INVdiaR2 and INVdiaR4 (CTC AGA GTA TTT TAT CCT CAT CTC). The cycling conditions for this reaction were an initial melting step (95°C, 1 min), followed by 35 cycles of 95°C for 1 min, 50°C for 45 secs, and 72°C for 45 secs.

Genotyping of *Dll4* alleles in mice

Genotyping of *Dll4* COIN mice was performed either by PCR or by Southern analysis. For PCR, reactions of 25μl total volume were set up, using ex-Taq DNA polymerase (TaKaRa) in the buffer and nucleotide concentrations recommended by its manufacturer, and containing 5% DMSO, and 0.5 μg of genomic DNA as the template. Primers 1407fwd (TGA GAG AGT TTC CTG GCG AAG TC) and 1407rev (TGC GGC ACA CAC AGA TGA AG) were used to score for the non-inverted COIN alleles (product size = 613 bp), whereas primers 1407INVfwd (CGG GTT ATT TTT GGC GTG GG) and 1407rev were used to score for the inverted COIN allele (product size = 599 bp). Each type of reaction was run separately. After an initial melting step (94°C, 4 min), 35 cycles (94°C, 1 min; 58°C, 45 secs; 72°C, 45 secs) of PCR were performed, followed by electrophoretic analysis.

For Southern analysis a probe encompassing part of intron 3 and exon 4 of *Dll4* (genomic coordinates 119,152,521 to 119,154,928 on Chromosome 2) was used to probe mouse genomic DNA restricted with *BglII*. The unmodified allele of *Dll4* results in 4.8 kb band, whereas $Dll4^{i3COIN}$ and $Dll4^{i3COIN\text{-}INV}$ result in 4.1 and 3.6 kb respectively.

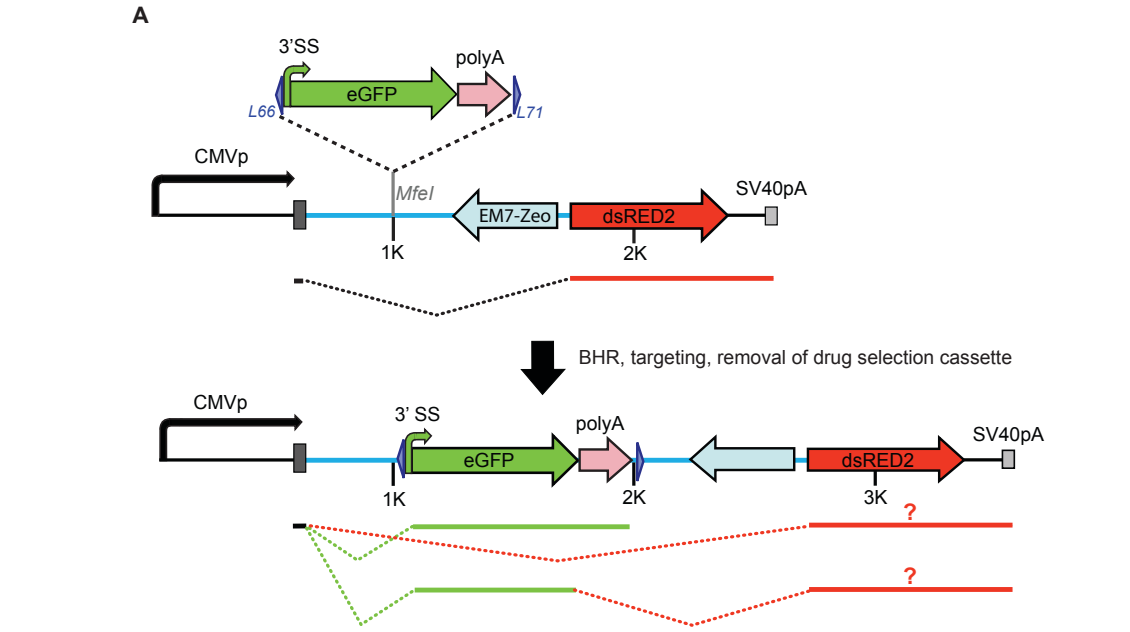
Quantitative RT-PCR

Probes and primers used for quantitative RT-PCR are listed in Dataset S2.

Statistical analysis of mouse crosses

Chi square and *P* values were calculated using the on-line calculator http://www.graphpad.com/quickcalcs/chisquared1.cfm.

Supplementary Data



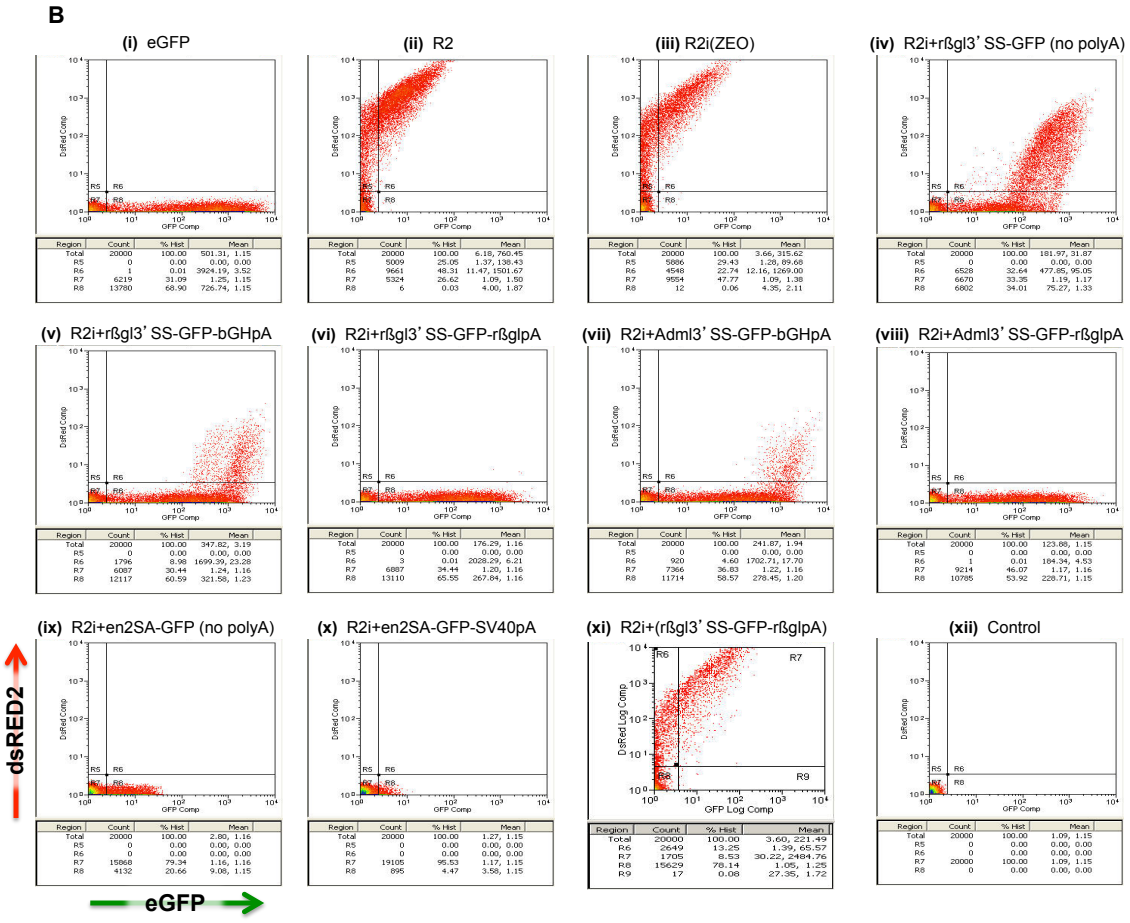


Fig. S1. COIN module optimization. (*A*) Schematic representation of the reporter mini-gene used to select optimal combinations of 3'SS-polyadenylation pairs for the gene-trap element of the COIN module. A CMV promoter (black curved arrow, CMVp) was used to drive the expression of dsRED2 (Clonetech) (red arrow). The ORF of dsRED2 was split into two exons using HBB2 intron 2 (light blue line, and Fig. S14), such that the ATG of dsRED2 resides in exon 1 (gray box), and hence shared with that of eGFP (green arrow), when the latter is introduced as a new exon into the *MfeI* site of HBB2 intron 2. Prior to insertion of different 3'SS-eGFP-pA cassettes, dsRED2 is expressed (upper panel). After insertion of different 3'SS-eGFP-pA cassettes either only eGFP, only dsRED2, or both will be expressed (lower panel). This outcome is determined by the specific combination of 3'SS-pA. An EM7 promoter-driven Zeo ORF was used during the construction of these vectors and is shown here (light blue arrow). The placement of the *lox66* (*L66*) and *lox71* (*L71*) sites flanking the 3'SS-eGFP-pA cassette is marked with blue arrowheads. (*B*) The 3'SS from either HBB2 intron 2 or adenovirus major late mRNA intron coupled to the pA region of HBB2 provide the most effective combinations for generating efficient COIN modules. Different combinations of 3'SS-pA pairs were tested for efficiency of splicing into eGFP, by transiently transfecting the corresponding expression vectors into HEK-293 cells (20) and two days later scoring the expression of dsRED2 and eGFP by Fluorescence Activated Cell Sorting (FACS). First, an intron-containing dsRED2 mini-gene was engineered by splitting the cDNA of dsRED2 in pDsRED2-N1 (Clonetech) into two exons using HBB2 intron 2 (*A*), and demonstrating that the intron-containing dsRED2 (iii) was expressed at the same level as its parental, single exon version (ii). Second, the requirement for a polyadenylation region was demonstrated by the comparing the expression of dsRED2 and eGFP after introducing a 3'SS-eGFP cassette into the intron of dsRED2; when no pA was included in the 3'SS-eGFP cassette (iv), both dsRED2 and eGFP were expressed (compare iv with i, ii, and iii). The most optimal combinations of 3'SS-pA pairs were those utilizing either the 3'SS from HBB2 (rßgl3'SS; GGG CCC CTC TGC TAA CCA TGT TCA TGC CTT CTT CTT TTT CCT ACA G) (21) or the adenovirus major late mRNA intron 3'SS (Adml3'SS; TAG GGC GCA GTA GTC CAG GGT TTC CTT

GAT GAT GTC ATA CTT ATC CTG TCC CTT TTT TTT CCA CAG) (4), coupled to the pA region of HBB2 (bglpA; Fig. S15) (3), as evidenced by the lack of expression of dsRED2 and a high expression level of eGFP (vi, viii). In contrast, the bovine growth hormone pA (bGHpA) (2) is not as efficient in terminating transcription when coupled either to the rßgl3'SS (v) or the Adml3'SS (vii) (22), indicating that the selection of the right 3'SS-polyA pair is indeed a critical parameter in the engineering of efficient gene-trap cassettes. Incidentally, gene-trap cassettes utilizing the 3'SS and upstream intron region from *engrailed 2* (en2SA; region from Chromosome 5, coordinates 28,494,951 to 28,496,659 in Build 58) either alone (ix) or in combination with the polyadenylation region (pA) from SV40 (SV40pA) (23) (x) are very effective in blocking expression of dsRED2 but were not chosen, firstly because the SV40pA contains polyadenylation signals both in the 'sense' and the 'antisense' strand, and secondly because the intron region from en2SA appears to reduce the expression of eGFP (compare ix and x with v through viii). Lastly, eGFP is not expressed if the 3'SS-eGFP-polyA cassette is placed in the antisense strand of the rabbit HBB2 intron, and more importantly, the presence of the 3'SS-eGFP-polyA cassette in the antisense strand of that intron does not adversely affect the expression of dsRED2 (xi). Panels showing the profile of eGFP (i), dsRED (ii), and untransfected cells (xii) are shown to provide standards for comparing the expression of eGFP and dsRED2 in the different constructs. The data presented in panel xi was derived from an independent experiment. Identical data to that shown here has been reproduced in COS7 cells and also partly *in vivo* using a COIN allele of *Gt(ROSA)26Sor* (Supplementary Appendix, Fig. S10), thereby indicating that these results are not restricted to one cell type, and that observations made in cell culture extend *in vivo*. The majority of the COIN alleles described in this paper make use of the rßgl3'SS-eGFP-rßglpA cassette for the COIN module, as the 3'SS and pA elements have coevolved. The expression vectors used in each panel were:

(i) pEGFP-N1 (eGFP; Clonetech)

(ii) pDsRED2-N1 (R2; Clonetech)

(iii) pDsRed2+rßgl IVS2/ZEO (R2i(ZEO))

(iv) pDsRed2+[rßgl.(L66)/GFP/L71] (R2i+rßg3'SS-GFP). Lacks pA.

(v) pR2i[rßgl.(L66)/GFP-bGHpA/L71] (R2i+rßg3'SS-GFP-bGHpA)

(vi) pR2i[rßgl.(L66)/GFP-ßglpA/L71 (R2i+rßg3'SS-GFP-bglpA)

(vii) pR2i[(L66)Adml_GFP-bGHpAL71] (R2i+Adml3'SS-GFP-bGHpA)

(viii) pR2i[(L66)_Adml_GFP_ßglpA-L71] (R2i+Adml3'SS-GFP_bglpA)

(ix) pR2i[(L66)-en2SAGFP-L71] (R2i+en2SA-GFP). Lacks pA.

(x) pR2i[(L66)-en2SAGFPsv40pA-L71] (R2i+en2SA-GFP-SV40pA)

(xi) pR2i[rßgl.(L71)/(GFP-ßglpA)L66] (R2i+(rßg3'SS-GFP-bglpA)

The 3'SS and polyA elements used in each cassette have also been listed in Table S11. The sequences of these vectors are available upon request.

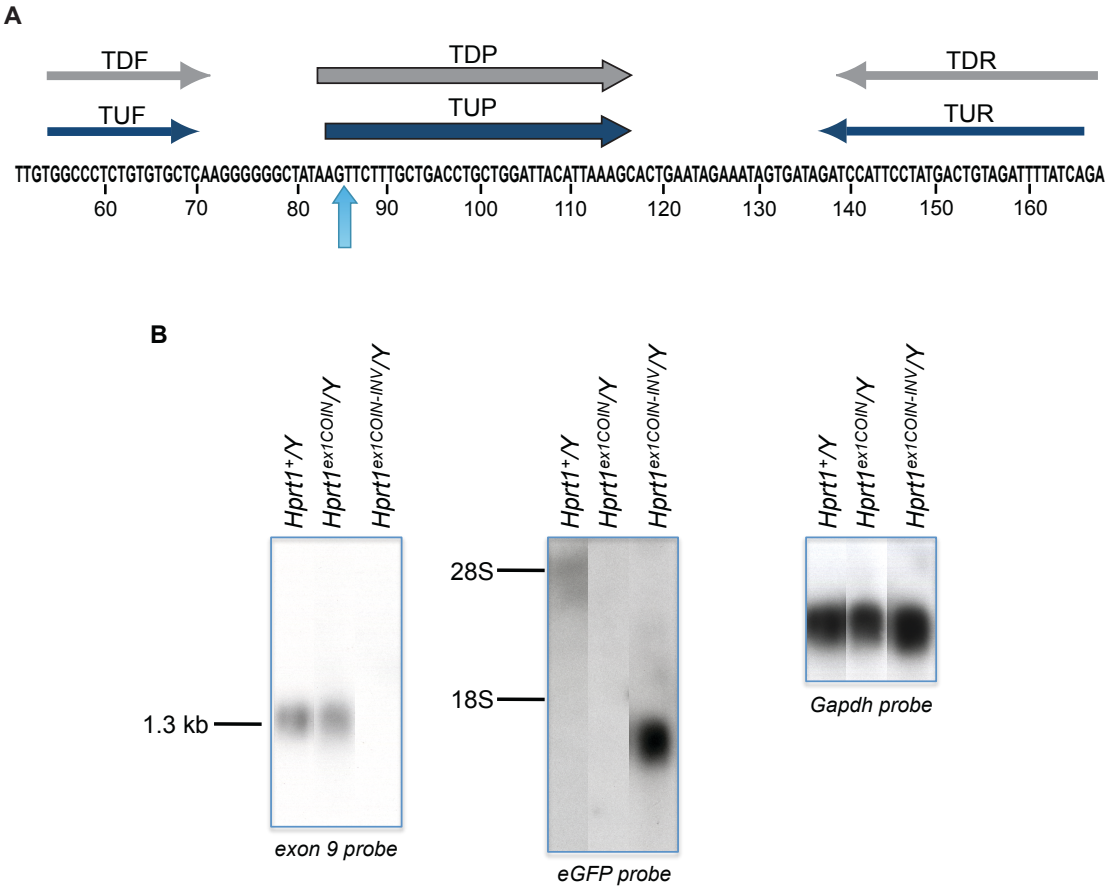


Fig. S2. (*A*) Sequence of *Hprt1*'s exon 3 around the insertion point of the COIN intron. The insertion point is after the AAG (after nt 85), and disrupts LOA probes TUP and TDP. Numbers under the sequence indicate the position of the corresponding nucleotide within exon 3. (*B*) The level and size of *Hprt1* mRNA is unaltered in *Hprt1*^{ex3COIN}*/Y* ES cells compared to wild type (*Hprt1*$^{+}$*/Y*). Total RNA was extracted from ES cells that are *Hprt1*$^{+}$*/Y*, *Hprt1*^{ex3COIN}*/Y*, *Hprt1*$^{ex3COIN\text{-}INV}$*/Y* and was subjected to Northern analysis. To visualize the *Hprt1* mRNA expressed in each ES cell line and compare it with *Gapdh* and *eGFP*, the same Northern blot was sequentially hybridized, being probed first with an exon 9 probe, then with an *eGFP* probe, and lastly with *Gapdh*. Hybridization to *Gapdh* (right panel) was used to provide an independent standard for the amount of RNA in each lane. *Hprt1*^{ex3COIN}*/Y* and *Hprt1*$^{+}$*/Y* ES cells express nearly identical levels of *Hprt1* mRNA, whereas

Hprt1$^{ex3COIN\text{-}INV}$*/Y* ES altogether lack *Hprt1* mRNA (left panel). In *Hprt1*$^{ex3COIN\text{-}INV}$*/Y* ES cells the *Hprt1* mRNA is replaced with one encoding for *eGFP* (center panel).

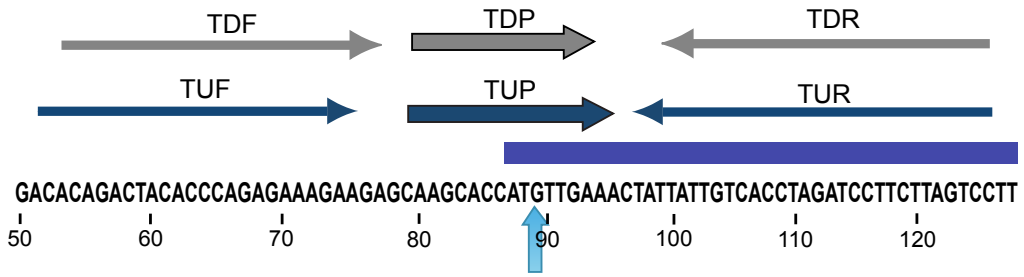


Fig. S3. Sequence of *Il2rg*'s exon 1 around the insertion point of the COIN intron. The insertion point is after the ATG (after nt 88), and disrupts LOA probes TUP and TDP. Numbers under the sequence indicate the position of the corresponding nucleotide within exon 1.

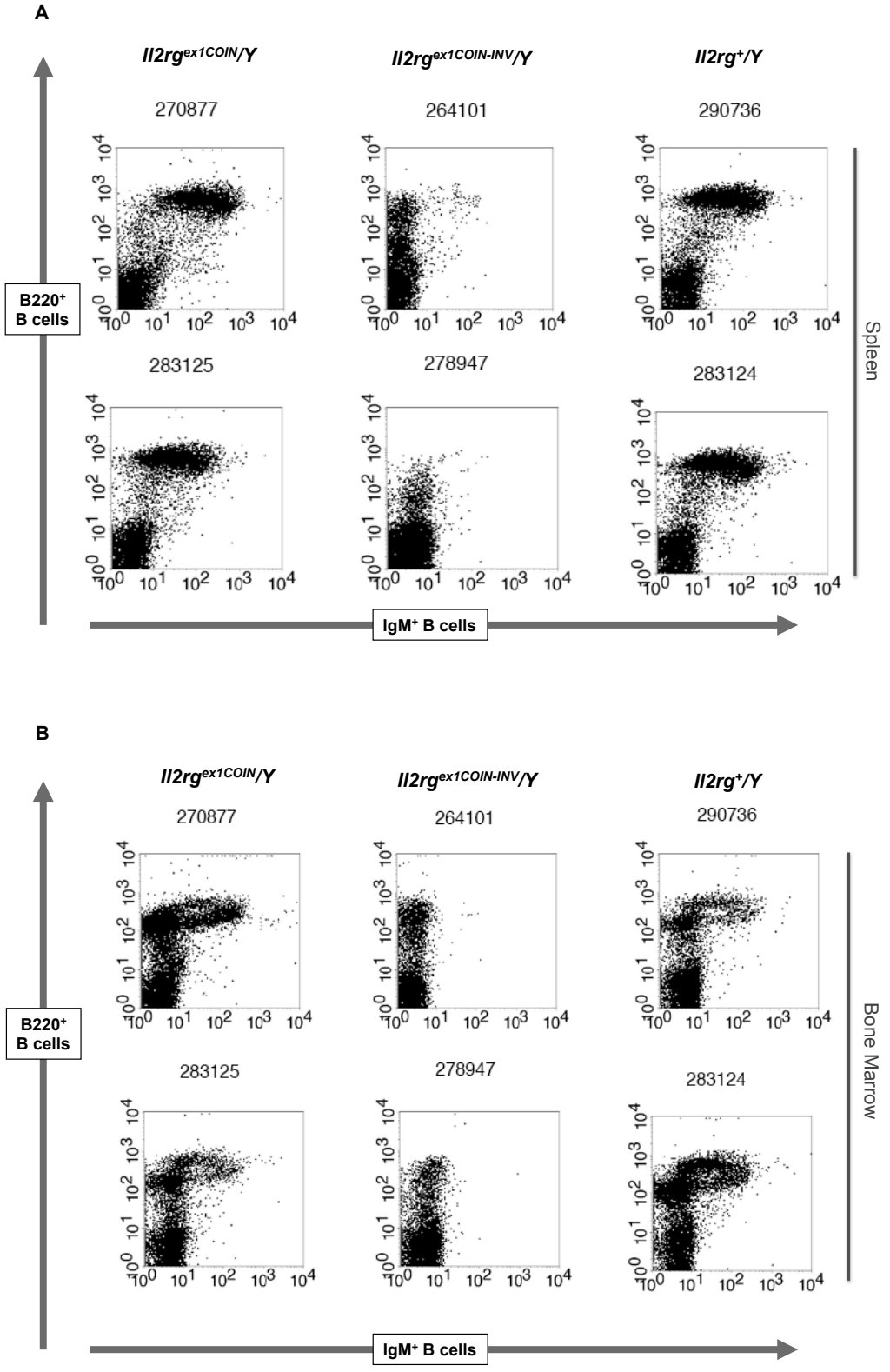

A
Il2rg^ex1COIN/Y
Il2rg^ex1COIN-INV/Y
Il2rg+/Y
270877
264101
290736
283125
278947
283124
B220+ B cells
IgM+ B cells
Spleen
B
Il2rg^ex1COIN/Y
Il2rg^ex1COIN-INV/Y
Il2rg+/Y
270877
264101
290736
283125
278947
283124
B220+ B cells
IgM+ B cells
Bone Marrow

Fig. S4. *Il2Rg*^{ex1COIN}*/Y* mice are wild type whereas *Il2Rg*^{ex1COIN}*/Y* mice are null. Splenocytes (*A*) and bone marrow cells (*B*) were analyzed for surface expression of IgM and B220. Spleen and bone marrow cell populations contain the same relative % of IgM$^+$, B220$^+$ cells when isolated from *Il2rg*$^+$*/Y* (290736, 283124) and *Il2rg*^{ex1COIN}*/Y* (270877, 283125) mice, whereas they are almost devoid of IgM$^+$, B220$^+$ cells when isolated from *Il2rg*$^{ex1COIN\text{-}INV}$*/Y* (278947, 283126) mice.

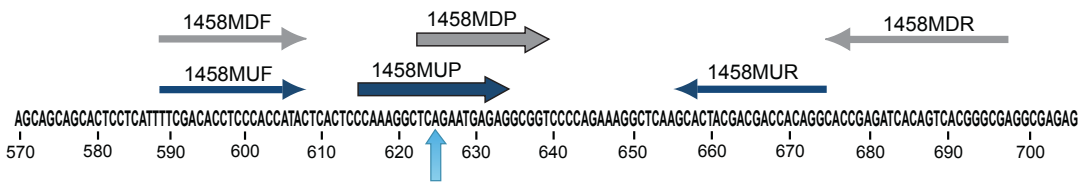


Fig. S5. Sequence of *Drosha*'s exon 4 around the insertion point of the COIN intron. The insertion point is after the CAG (after nt 624), and disrupts LOA probes TUP and TDP. Numbers under the sequence indicate the position of the corresponding nucleotide within exon 4.

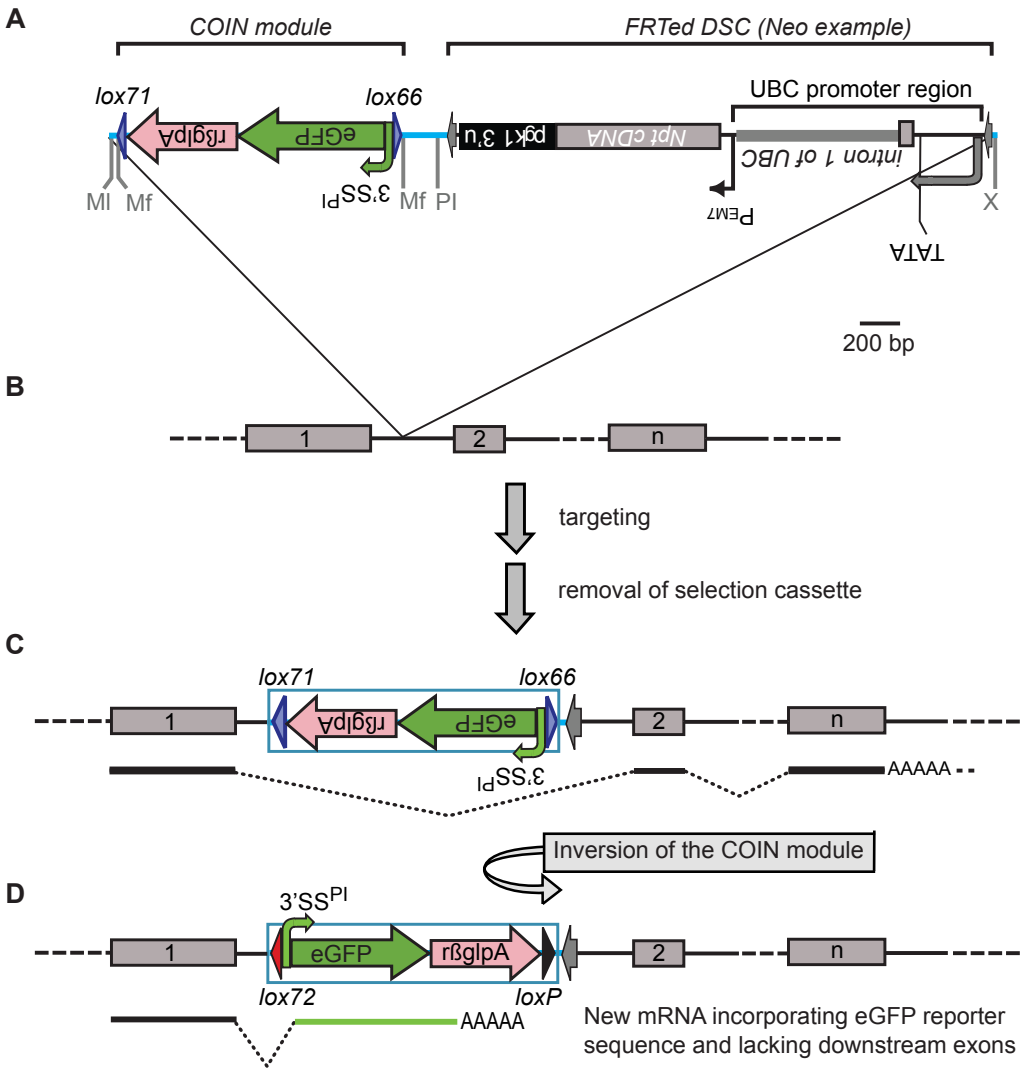


Fig. S6. COIN module insertion into native introns. The COIN module is identical to that described in Figure 1, with the exception that the intron regions upstream of the COIN module and downstream of the DSC are not incorporated. All other parameters are as described in Figure 1.

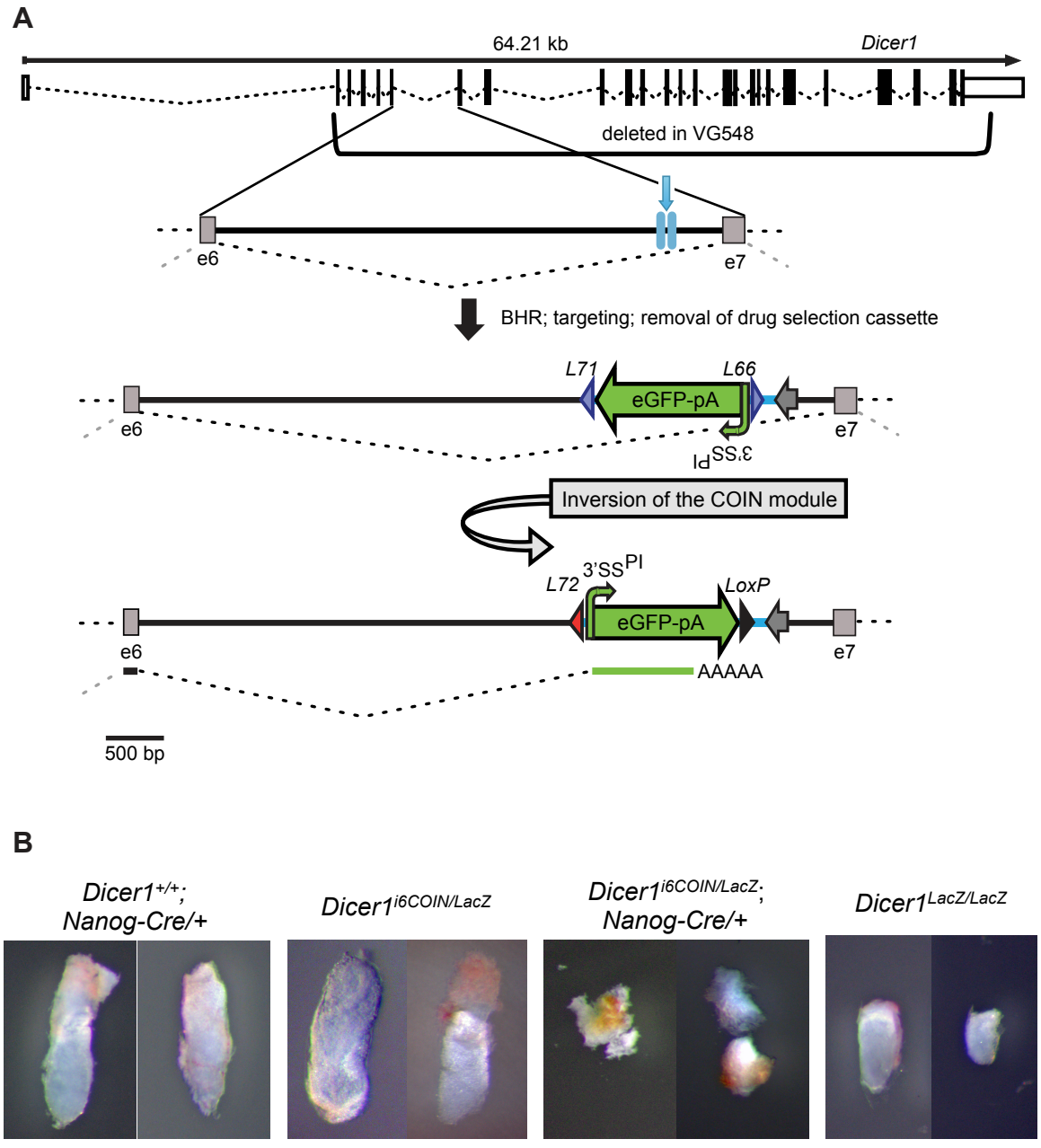


Fig. S7. *Dicer1*$^{i6COIN\text{-}INV}$ is a null allele. (*A*) Schematic of *Dicer1*^{i6COIN} allele. The exon-intron region of *Dicer1* adapted from Ensembl.org. The COIN module was inserted into intron 6 (replacing nucleotides 105,962,727 to 105,962,772 of Chromosome 12). Prior to inversion, the *Dicer1*^{i6COIN} allele generates a normal message as the COIN intron is spliced out. After inversion, the COIN module becomes the last functional exon of the modified gene, abrogating transcription of the downstream exons, and resulting in a functional null allele incorporating eGFP. Naming

conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bar, 500 bp. (*B*) The *Dicer1*$^{i6COIN\text{-}INV}$ allele corresponds to a null allele. Embryos from a *Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$*; Nanog-Cre* or a *Dicer1*$^{LacZ/+}$ x *Dicer1*$^{LacZ/+}$ cross were collected at E7.5, visualized by light microscopy and genotyped by PCR. The *Dicer1*$^{i6COIN/LacZ}$; *Nanog-Cre/+* embryos phenocopy the *Dicer1*$^{LacZ/LacZ}$ embryos as well as the corresponding published *Dicer1-null* line (24).

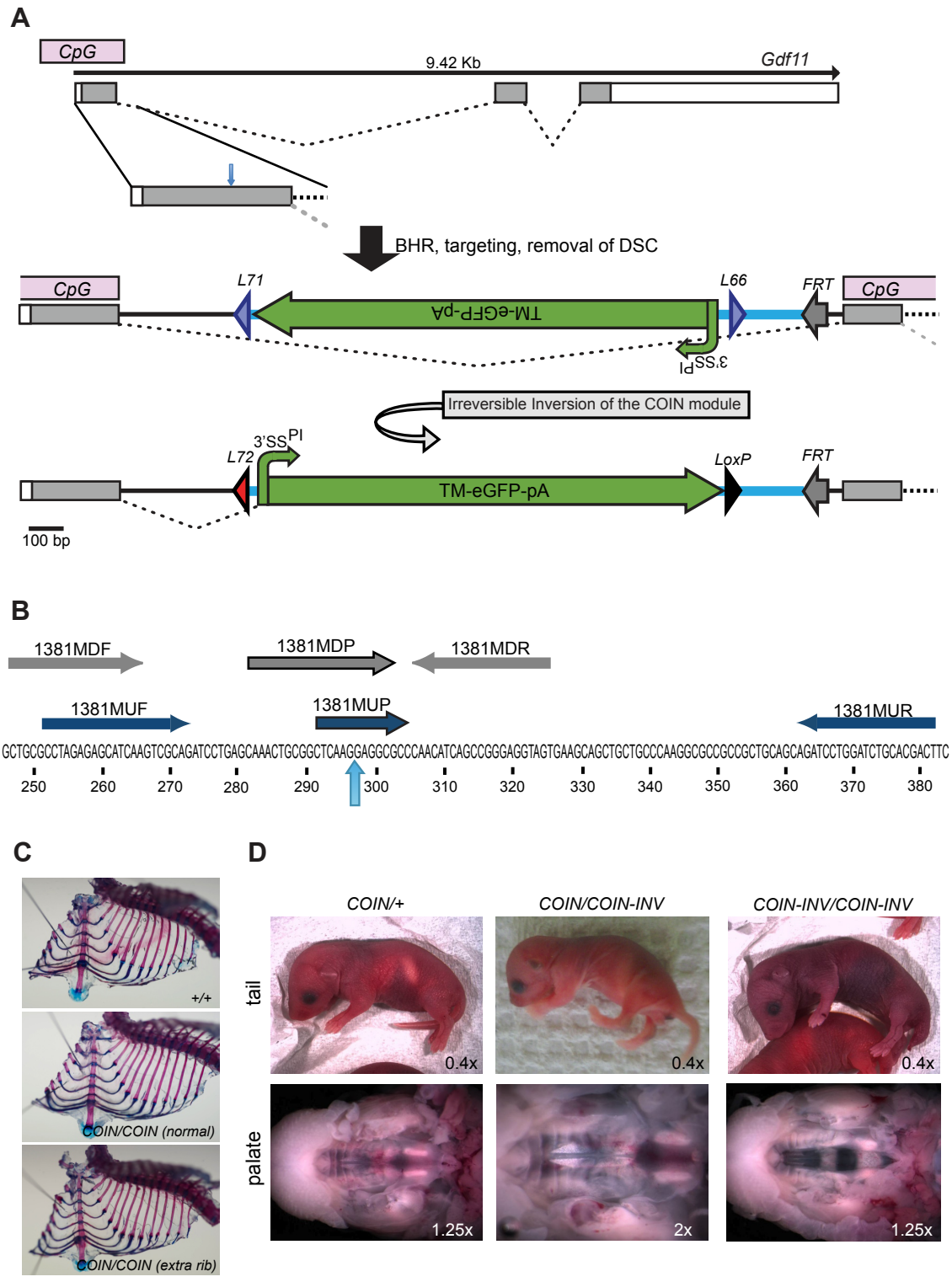
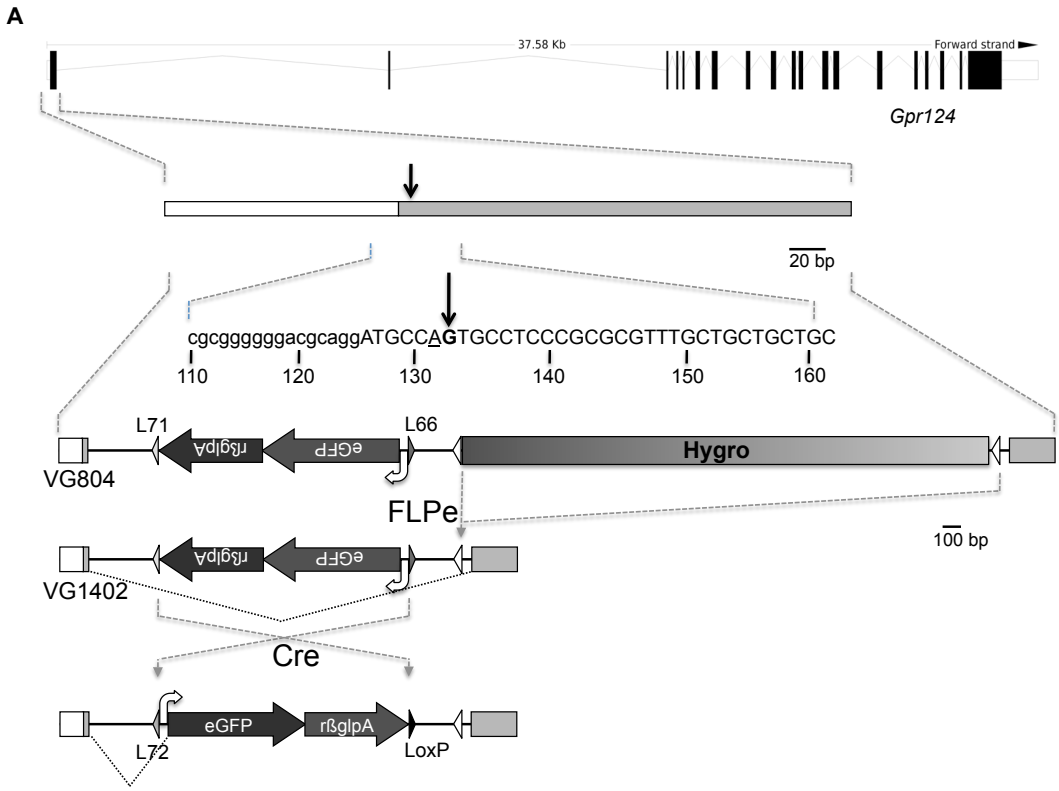


Fig. S8. *Gdf11*^{ex1COIN} is wild type whereas *Gdf11*$^{ex1COIN\text{-}INV}$ is a null allele. (*A*) Schematic of *Gdf11*^{ex1COIN} allele. The exon-intron region of *Gdf11* adapted from

Ensembl.org. The COIN module was inserted into exon 1, disrupting a CpG island. Prior to inversion, the *Gdf11*^{ex1COIN} allele generates a normal message as the COIN intron is spliced out. After inversion, the COIN module becomes the last functional exon of the modified gene, abrogating transcription of the downstream exons, and resulting in a functional null allele incorporating TMeGFP. Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bar, 100 bp. (*B*) Sequence of *Gdf11*'s exon 1 around the insertion point of the COIN intron. The insertion point is after the AAG (after nt 297), and disrupts LOA probes TUP and TDP. Numbers under the sequence indicate the position of the corresponding nucleotide within exon 1. (*C*) *Gdf11*$^{ex1COIN/ex1COIN}$ mice occasionally present with an extra rib. The rib cages of four month old *Gdf11*$^{+/+}$ (*+/+*) and *Gdf11*$^{ex1COIN/ex1COIN}$ (*COIN/COIN*) mice were evaluated histologically using Alizarin Red and Alcian Blue staining, which stain for proteoglycans (blue) and collagen-associated proteoglycans (fuchsia), to visualize cartilage and calcified bone respectively (25). 6 out of 19 *Gdf11*$^{ex1COIN/ex1COIN}$ mice display an additional thoracic rib (right panel; Table S12), whereas 89 out of 90 Gdf11 heterozygous-null mice have the extra thoracic rib (26). This indicates that the *Gdf11*^{ex1COIN} allele is slightly hypomorphic. (*D*) *Gdf11*$^{ex1COIN\text{-}INV/ex1COIN\text{-}INV}$ mice phenocopy the *Gdf11* null mice. P1 pups born to a *Gdf11*$^{ex1COIN/+}$ x *Gdf11*$^{ex1COIN\text{-}INV/+}$*; Nanog-Cre/+* cross were scored for anatomic phenotypes displayed by *Gdf11* null mice, such as short tail and cleft palate (26). Whereas *Gdf11*$^{ex1COIN/+}$ (*COIN/+*) and *Gdf11*$^{ex1COIN/ex1COIN\text{-}INV}$ (*COIN/COIN-INV*) mice display no phenotypic abnormalities (left and middle panels), *Gdf11*$^{ex1COIN\text{-}INV/ex1COIN\text{-}INV}$ (*COIN-INV/COIN-INV*) mice display both a shortened tail and a cleft palate (right panel; Table S13).


A
37.58 Kb
Forward strand
Gpr124
20 bp
cgcgggggggacgcaggATGCCAGTGCCTCCCGCGCGTTTGCTGCTGCTGC
110
120
130
140
150
160
L71
rßglpA
eGFP
L66
Hygro
VG804
FLPe
100 bp
VG1402
Cre
eGFP
rßglpA
L72
LoxP

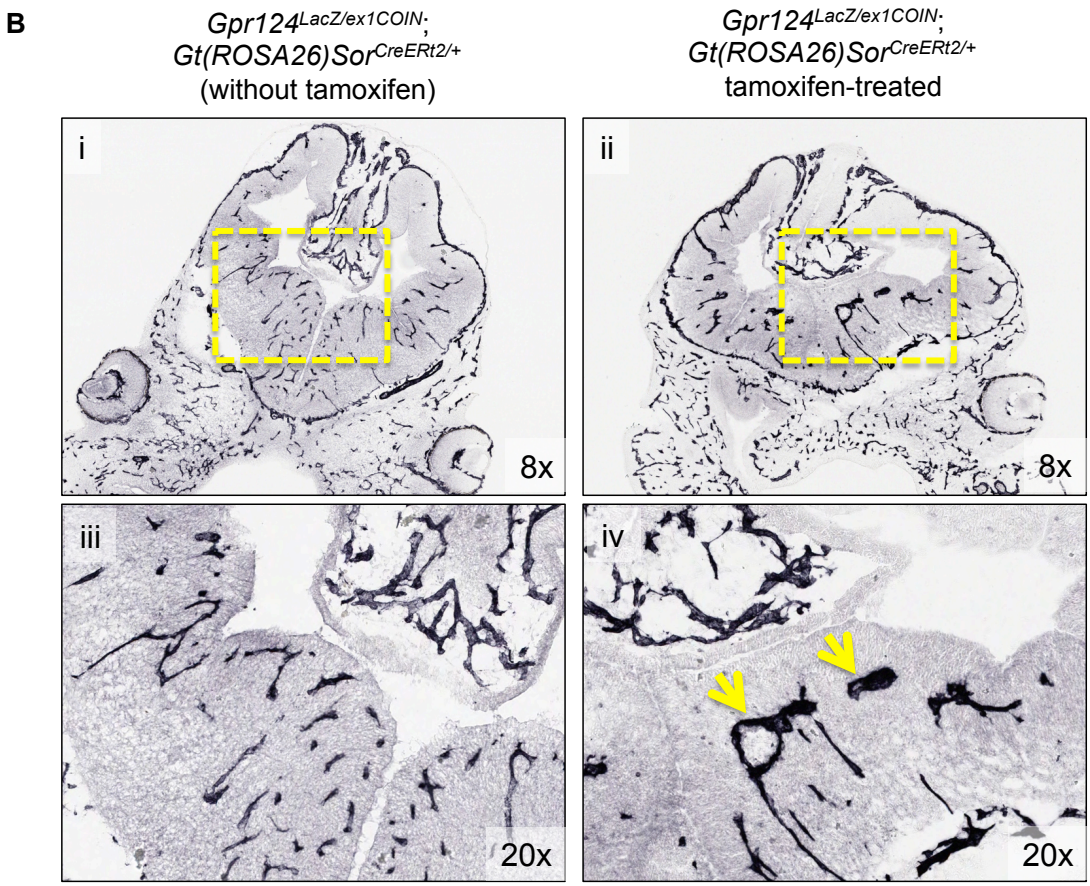
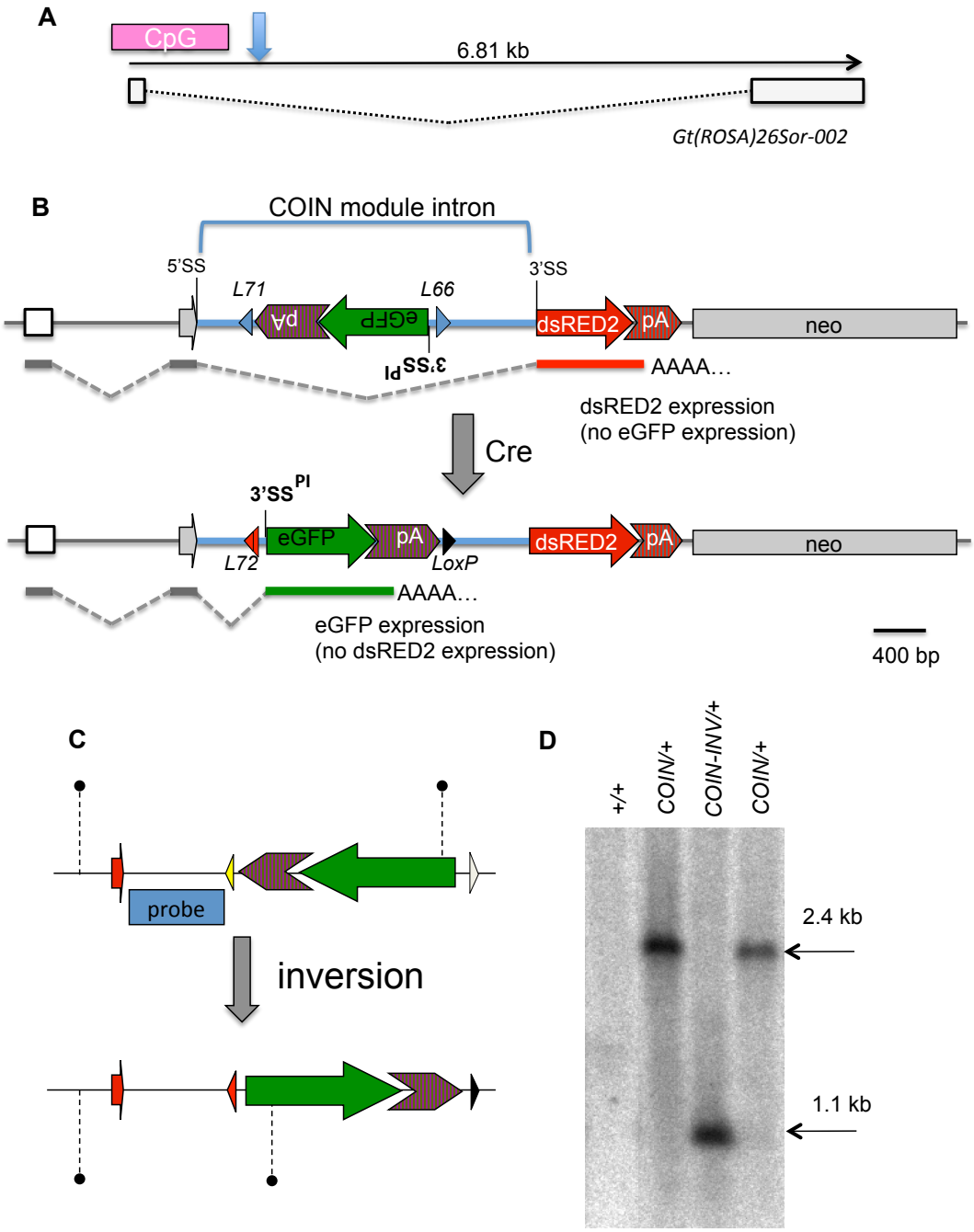

B
Gpr124LacZ/ex1COIN; Gt(ROSA26)SorCreERt2/+ (without tamoxifen)
Gpr124LacZ/ex1COIN; Gt(ROSA26)SorCreERt2/+ tamoxifen-treated
i
ii
8x
8x
iii
iv
20x
20x

Fig. S9. Gpr124$^{Laz/ex1COIN}$ embryos are phenotypically wild type whereas Gpr124$^{Laz/ex1COIN\text{-}INV}$ embryos phenocopy homozygous-null Gpr124 embryos. (*A*) Design of *Gpr124*^{ex1COIN} allele. The exon/intron structure of *Gpr124* gene is shown at top (adapted from Ensembl.org). Exon 1 (371 bp) is expanded; the 5'UTR is shown in white, whereas the coding sequence is gray. The vertical arrow marks the point of insertion of the COIN intron. DNA sequence in small case indicates 5' UTR sequence, whereas capital case denotes the coding sequence. The COIN intron was inserted after the G shown in bold. A silent C to A mutation, engineered to conform to the 5' splice site consensus, is underlined. After insertion of the COIN intron, exon 1 is divided into two exons of 133 bp 5' and 238 bp 3' of the intron, respectively. Expression of *Hygro* is driven by two promoters, $P_{hUB(C)}$ (for expression in mammalian cells) and P_{EM7} (for expression in *E. coli*), and it is followed by the mouse PGK polyadenylation signal and associated sequence (27). Post-targeting, the COIN allele VG804 is matured to the final COIN allele by treatment with FLPe, to generate allele VG1402. Exposure to Cre results in inversion of the COIN module into the sense strand, resulting in an allele encoding for a truncated mRNA, comprised of the 5' half of exon 1 and the eGFP exon of the COIN module. (*B*) Tamoxifen-induced global deletion of *Gpr124* at E8 indicates a requirement for Gpr124 expression shortly before and/or during the period of angiogenic sprouting into the brain. The brain of *Gpr124*$^{Laz/ex1COIN}$; *Gt(ROSA26)Sor*$^{CreERT2/+}$ double-mutant embryos develop normally (i, iii). In the presence of tamoxifen, dosed at E8, E12.5 *Gpr124*$^{Laz/ex1COIN}$; *Gt(ROSA26)Sor*$^{CreERT2/+}$ embryos display vascular abnormalities in the ventral forebrain at E12.5 (ii, iv) [yellow arrowheads in (iv)] that were similar to, though less severe than, those seen in the *Gpr124*$^{LacZ/LacZ}$ embryos (15). Boxes in (i) and (ii) show areas enlarged in (iii) and (iv), respectively. The magnification is indicated on each panel. The vasculature is visualized by PECAM staining (28).


A
CpG
6.81 kb
Gt(ROSA)26Sor-002
B
COIN module intron
5'SS
3'SS
L71
L66
pA
eGFP
dsRED2
pA
neo
AAAA…
dsRED2 expression
(no eGFP expression)
Cre
3'SS PI
eGFP
pA
L72
LoxP
dsRED2
pA
neo
AAAA…
eGFP expression
(no dsRED2 expression)
400 bp
C
probe
inversion
D
+/+
COIN/+
COIN-INV/+
COIN/+
2.4 kb
1.1 kb

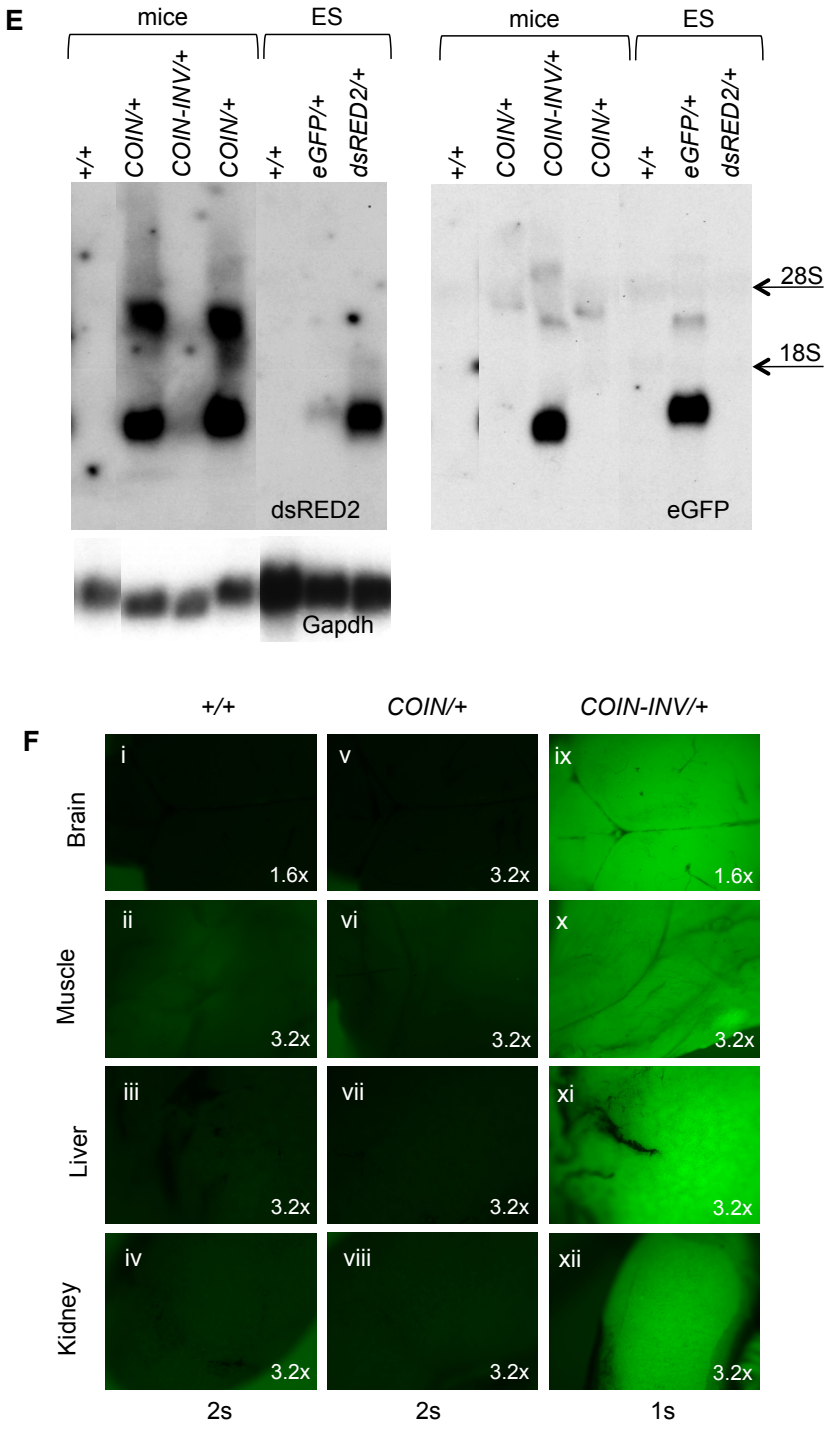
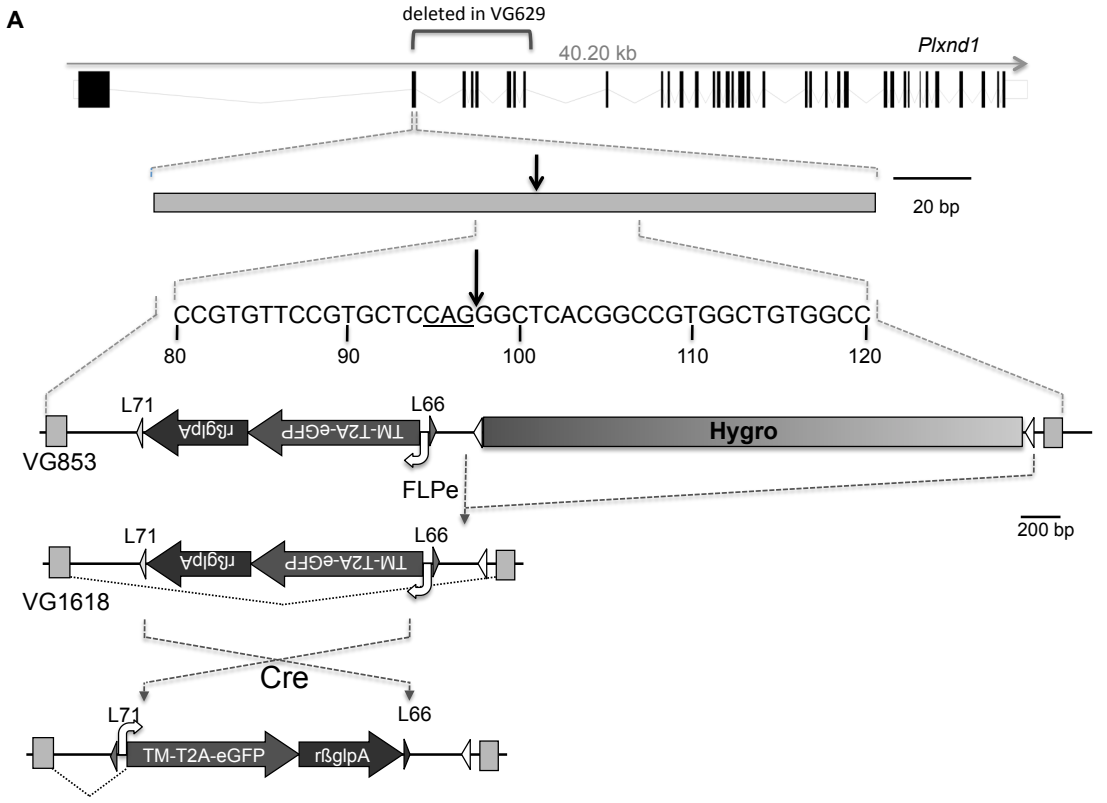
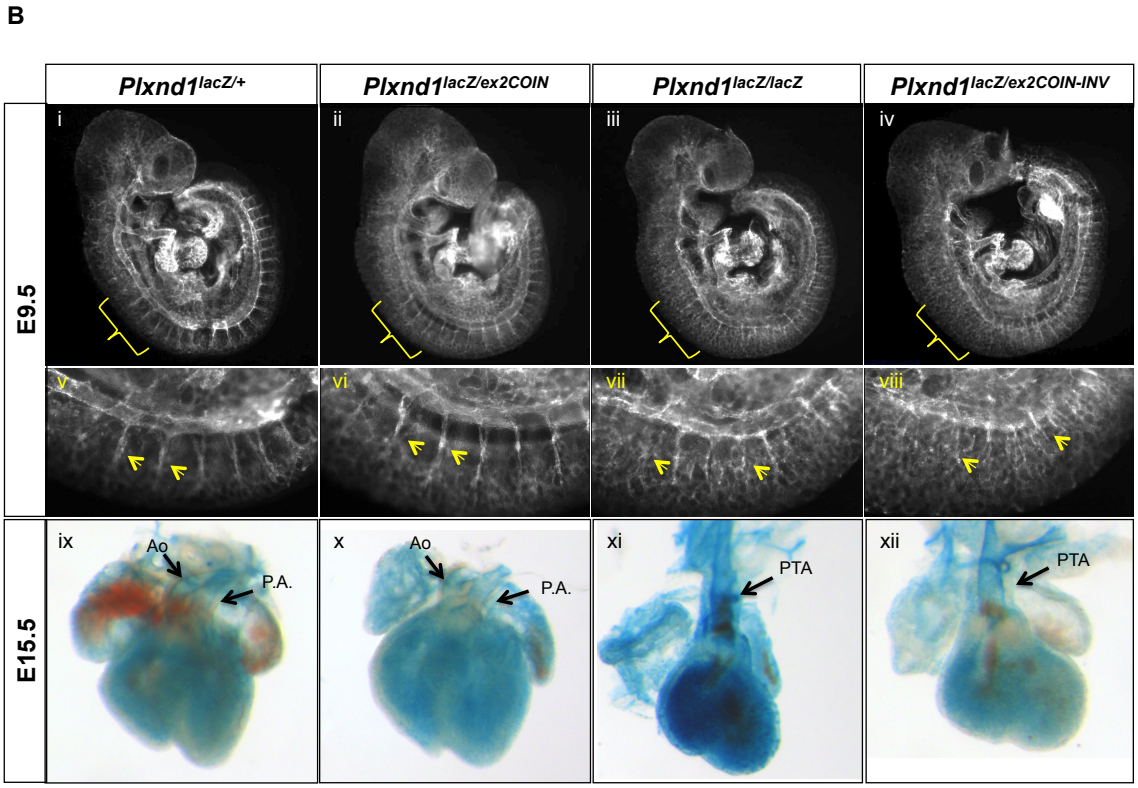


Fig. S10. *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ allele shows application of COIN to a non-coding RNA and demonstrates the usefulness of the reporter. (*A*) The exon-intron

structure of *Gt(ROSA26)Sor* was adapted from Ensembl.org. (*B*) Schematic of *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ allele. A reporter comprised of the 3' spice region from adenovirus major late transcript (22) followed by a dsRED2 ORF with an embedded COIN module was inserted at the *XbaI* site within the first intron of *Gt(ROSA26)Sor*, replacing nucleotides 113026005 to 113026068 of Chromosome 6. The first coding exon (gray arrow) of this reporter is shared between dsRED2 and the eGFP encoded by the COIN module. Prior to inversion, the *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ allele generates a message encoding dsRED2, as the COIN intron is spliced out. After inversion, the COIN module becomes the last functional exon of the modified gene, abrogating transcription of dsRED2, and replacing it with that of eGFP. Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2, except that the pA after dsRED2 is that from bovine growth hormone (bGH pA). Scale bar, 400 bp. (*C*), (*D*) Southern analysis to determine inversion of the COIN module in the *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ allele. Restriction with *PspOMI* and hybridization to a probe comprised of the sequence defined by the 5' splice site and the Lox71 site of the COIN intron yields a ~2.4 kb fragment for the *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ allele (*COIN/+*) and a 1.1 kb fragment for the inverted COIN allele, *Gt(ROSA26)Sor*$^{i1COIN\text{-}INV(eGFP)}$ (*COIN-INV/+*). (*E*) Northern analysis demonstrates robustness of COIN module in terminating transcription. dsRED2 expression is detected only in *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)/+}$ (*COIN/+*) ES cells and mice whereas it is completely absent from *Gt(ROSA26)Sor*$^{i1COIN\text{-}INV(eGFP)/+}$ (*COIN-INV/+*) ES cells and mice. RNA isolated from *Gt(ROSA26)Sor*$^{eGFP/+}$ and *Gt(ROSA26)Sor*$^{dsRED2/+}$ ES cells provide positive controls for the hybridization. Note that the longer mRNA detected for dsRED2 reflects readthrough transcription past the bGH pA; no such message is detected for eGFP, as the COIN module utilizes the efficient 3'SS^{PI}-pA pair, selected as described on Figure S1. (*F*) Whole mount visualization for eGFP fluorescence demonstrates utility of the reporter embedded in the COIN module. eGPF expression is detected only in tissues isolated from *Gt(ROSA26)Sor*$^{i1COIN\text{-}INV(eGFP)/+}$ (*COIN-INV/+*) mice (panels ix to xii).


A
deleted in VG629
40.20 kb
Plxnd1
20 bp
CCGTGTTCCGTGCTCCAGGGCTCACGGCCGTGGCTGTGGCC
80
90
100
110
120
L71
L66
rßglpA
TM-T2A-eGFP
Hygro
VG853
FLPe
200 bp
VG1618
Cre
TM-T2A-eGFP
rßglpA


B
Plxnd1lacZ/+
Plxnd1lacZ/ex2COIN
Plxnd1lacZ/lacZ
Plxnd1lacZ/ex2COIN-INV
E9.5
E15.5
Ao
P.A.
PTA

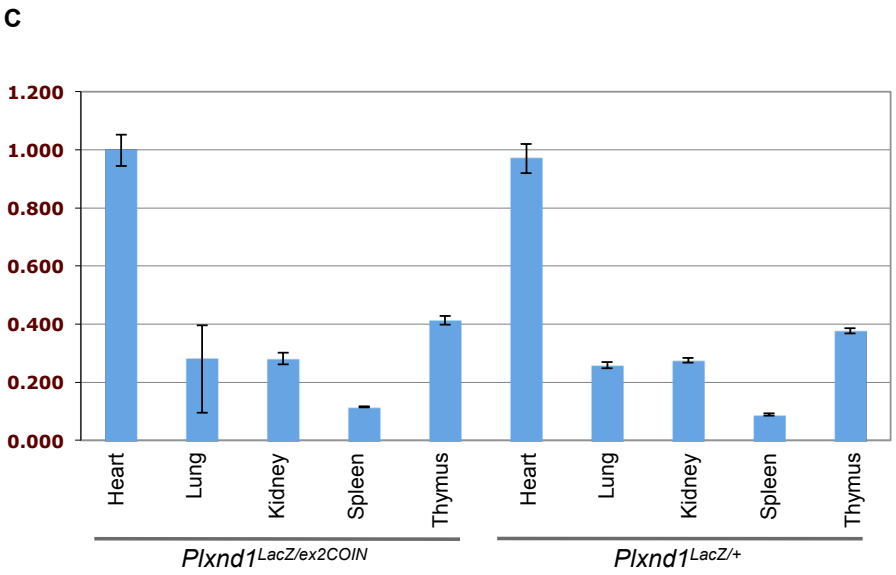


Fig. S11. (*A*) Schematic of *Plxnd1*^{ex2COIN} allele. The exon-intron region of *Plxnd1* adapted from Ensembl.org. The COIN module was inserted into Plxnd1's exon 2, after the CAG (underlined), i.e. after nucleotide 115930772 on Chromosome 6. Prior to inversion, the *Plxnd1*^{ex2COIN} allele (VG1618) generates a normal message as the COIN module is antisense to *Plxnd1*, and the COIN intron is spliced out. After inversion, the COIN module becomes the last functional exon of the modified gene, abrogating transcription of the downstream exons, and resulting in a functional null allele incorporating eGFP. Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bars as noted. (*B*) The *Plxnd1*$^{ex2COIN\text{-}INV}$ allele (i.e. the post-inversion allele) recapitulates the *Plxnd1* null phenotype. (i, ii, iii, iv) Immunofluorescent detection of the vascular endothelial marker Pecam1 (CD31) in whole-mount embryos shows the development of the vascular network at mid-gestation (E9.5). Middle panels (v, vi, vii, viii) are higher magnification images of bracketed regions depicting the segmented intersomitic vessels opposite, and immediately rostral to, the developing forelimb bud. By comparison to *Plxnd1*$^{lacZ/+}$ (i) and *Plxnd1*$^{lacZ/ex2COIN}$ (ii) embryos, which exhibit well-patterned vessels that emanate from the dorsal aorta into the intersomitic spaces, *Plxnd1*$^{lacZ/lacZ}$ show a severe reduction in intersomitic vessel formation and aberrant vascular sprouting within the intersomitic spaces (iii, and vi – arrowheads).

Inversion of the COIN allele results in a similar vascular phenotype (iv, and viii – arrowheads). (ix, x, xi, xii) Whole mount analysis of embryonic hearts that were dissected at E15.5 and stained for β-galactosidase activity. While both *Plxnd1*$^{lacZ/+}$ (ix) and *Plxnd1*$^{lacZ/ex2COIN}$ (x) embryos exhibit full septation of the outflow track, resulting in separation of the aorta (Ao) and pulmonary artery (P.A., arrows), *Plxnd1*$^{lacZ/lacZ}$ (xi) and *Plxnd1*$^{lacZ/ex2COIN\text{-}INV}$ (xii) embryos develop persistent truncus arteriosis (PTA, arrows). (*C*) The COIN module does not alter the expression of *Plxnd1*. Quantitative RT-PCR of different mouse tissues collected from age- and gender-matched littermates (n = 4). Mouse tissues (heart, lung, kidney, spleen and thymus) were collected and stored in RNAlater and RNAs were extracted using standard protocols. Quantative RT-PCR analysis was carried out with the TaqMan method with ABI Assay ID Mm01184367_m1* probe set (Applied Biosystems, Inc.) to specifically detect the expression of mouse PlexinD1 gene. Expression of mouse beta actin gene (*Actb*) was used as house keeping gene control to normalize the relative expression levels of *Plxnd1* mRNA. There is no significant difference in the expression level of *Plxnd1* mRNA as expressed from the *Plxn1*^{ex2COIN} allele compared to the wild type allele (*Plxnd1*$^{+}$).

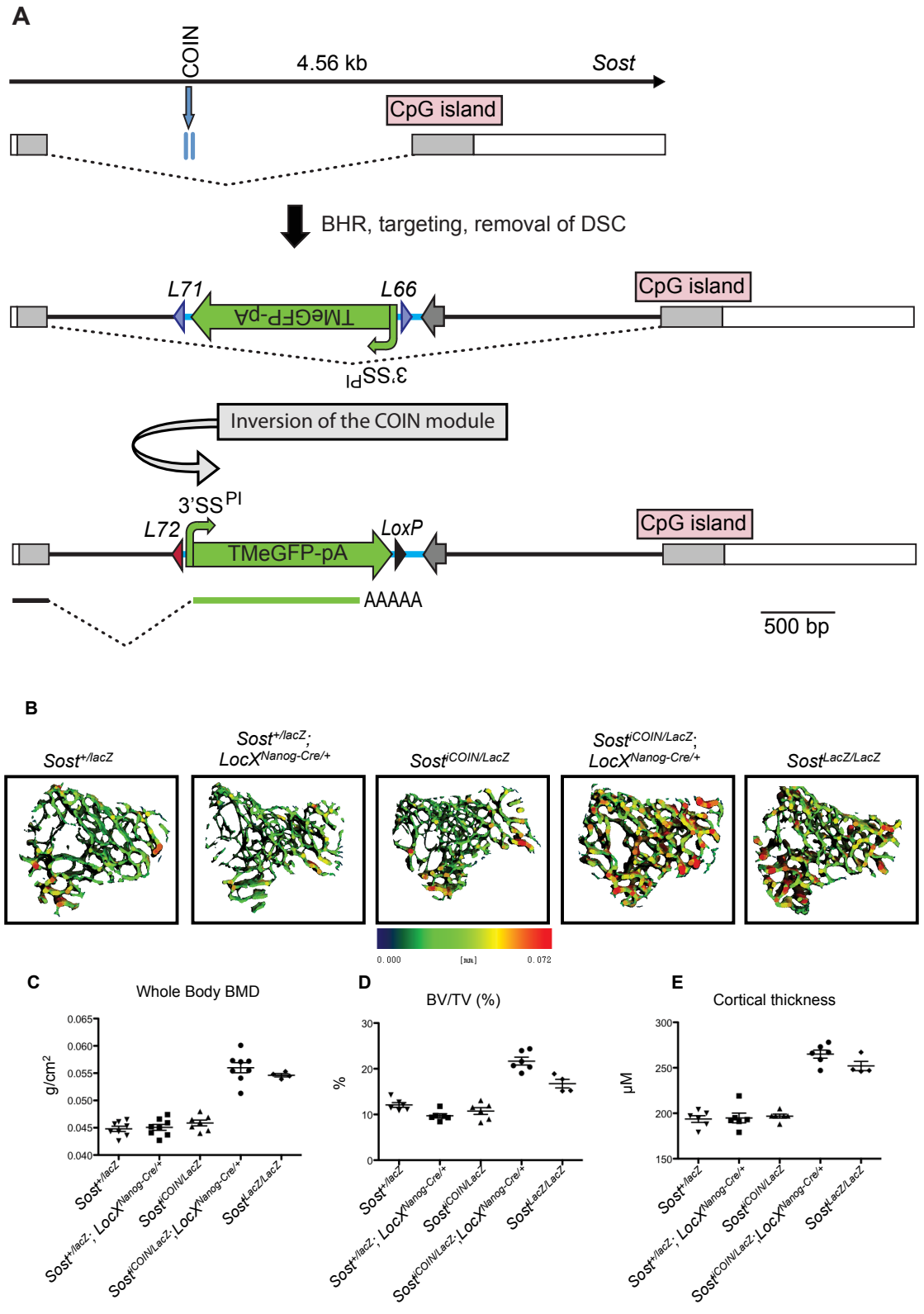


Fig. S12. *Sost*^{iCOIN} is wild type whereas *Sost*$^{iCOIN\text{-}INV}$ is a null allele. (*A*) Schematic of *Sost*^{iCOIN} allele. The exon-intron region of *Sost* adapted from Ensembl.org. The COIN

module was inserted into *Sost*'s single intron (replacing nucleotides 101,827,086 to 101,827,135 of Chromosome 11). Prior to inversion, the $Sost^{iCOIN}$ allele generates a normal message as the COIN module is spliced out. After inversion, the COIN module becomes the last functional exon of the modified gene, abrogating transcription of the downstream exons, and resulting in a functional null allele incorporating eGFP. Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bar, 500 bp. (*B*), (*C*), (*D*), (*E*) $Sost^{iCOIN\text{-}INV/LacZ}$ mice exhibit high skeletal mice, phenocopying $Sost^{LacZ/LacZ}$ mice, as revealed by analysis of the trabecular structure of the proximal tibial metaphysis from two month-old female mice using μCT. The tibial bones of $Sost^{iCOIN\text{-}INV/LacZ}$ (*iCOIN/LacZ; Nanog-Cre*/+) mice exhibit high bone mass, as do the $Sost^{LacZ/LacZ}$ (*LacZ/LacZ*) mice. Trabecular bone thickness of the proximal tibial metaphysis is visualized by pseudocoloring (*B*). The increased bone mass observed in $Sost^{iCOIN\text{-}INV/LacZ}$ and $Sost^{LacZ/LacZ}$ mice is reflected in the whole body bone mineral density calculated by DEXA (*C*), and translates to nearly a doubling of the tibial trabecular bone fraction (BV/TV, calculated by μCT) (*D*), as well as an approximately 37% increase in cortical thickness of tibial mid-diaphysis (calculated by μCT) (*E*). These phenotypes coincide with a published knockout line of *Sost* (29). A detailed phenotypic analysis of the $Sost^{iCOIN}$ will appear elsewhere (Collette et al, in preparation).

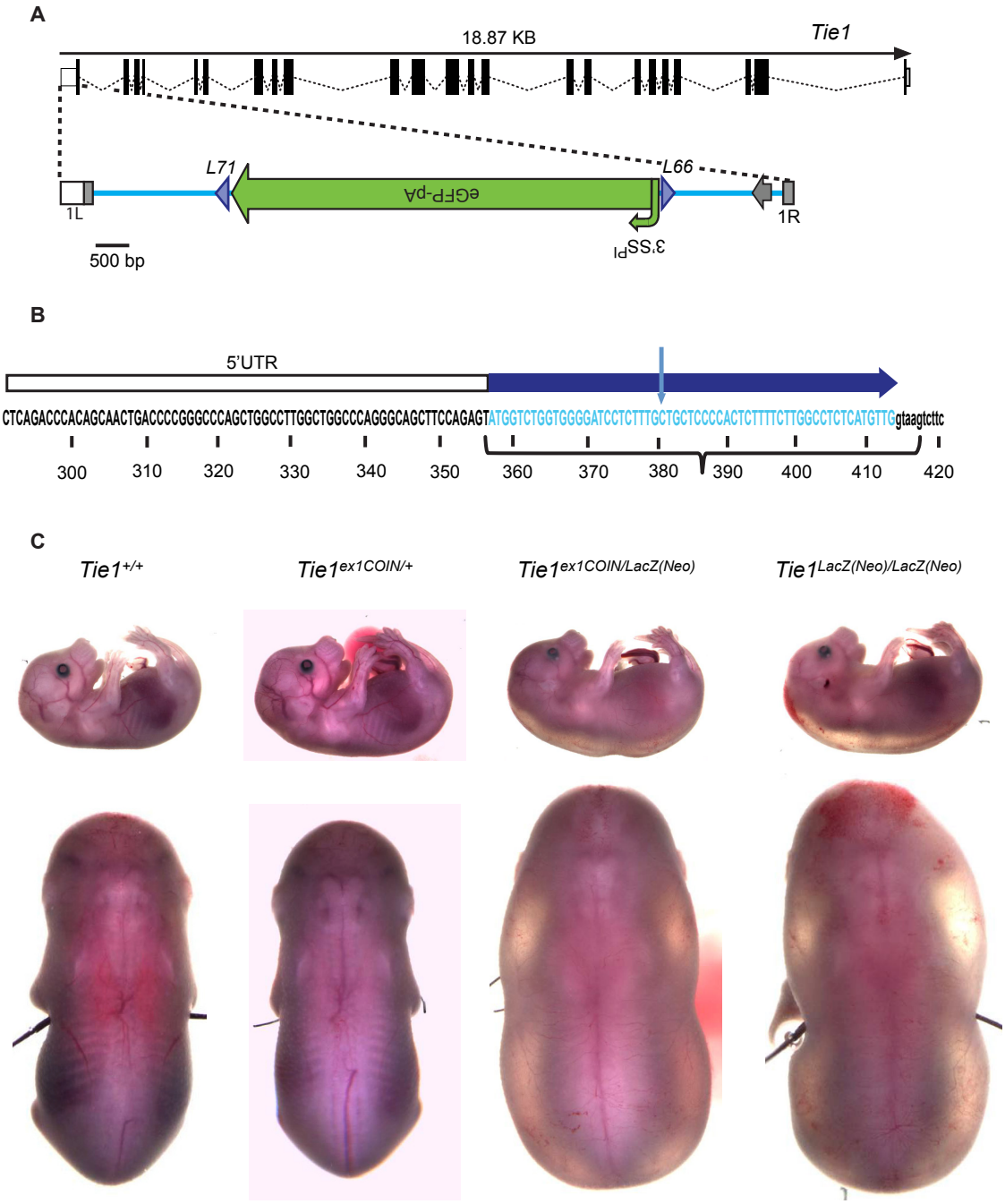


Fig. S13. *Tie1*^{ex1COIN} is wild type whereas *Tie1*$^{ex1COIN\text{-}INV}$ is a null allele. (*A*) Schematic of *Tie1*^{ex1COIN} allele. The exon-intron region of *Tie1* adapted from Ensembl.org. The COIN module was inserted into *Tie1*'s exon 1 (exon ENSMUSE00000756638 in isoform Tie1-001, ENSMUST00000047421; Chromosome 4) splitting exon 1 into exon 1L (380 nt) and 1R (34 bp). Note that exon 1R is relatively short. (*B*) The

COIN intron was inserted after the TTG, i.e. after nt 380 of exon 1. A vertical arrow marks the insertion point. 5'UTR sequence appears in black lettering whereas protein-coding sequencing appears in light blue lettering, both in upper case. Intronic sequencing at the start of intron 1 is shown in lowercase. The bracket underneath the sequence denotes the region of *Tie1* that was replaced with LacZ/Neo to generate a null allele of *Tie1* (VG1222). Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bar, 500 bp. (*C*) $Tie1^{ex1COIN\text{-}INV/LacZ}$ E15.5 embryos display the same phenotype as $Tie1^{LacZ/LacZ}$. E10.5 embryos were collected from a $Tie1^{+/LacZ}$ x $Tie1^{ex1COIN\text{-}INV/+}$ cross and compared to E10.5 embryos collected from a $Tie1^{+/LacZ}$ x $Tie1^{ex1COIN/+}$ cross. Whereas $Tie1^{ex1COIN/+}$ embryos are normal (2nd panel), $Tie1^{ex1COIN\text{-}INV/LacZ}$ embryos (3rd panel) display the same edema phenotype as $Tie1^{LacZ/LacZ}$ embryos (4th panel), and also as that reported for an independently developed *Tie1* knockout line (30). No phenotype has been associated with the $Tie1^{ex1COIN/LacZ}$ genotype (Table S10).

```
        10          20          30          40          50          60          70          80          90
GTGAGTTTGG GGACCCTTGA TTGTTCTTTC TTTTTCGCTA TTGTAAAATT CATGTTATAT GGAGGGGGCA AAGTTTTCAG GGTGTTGTTT
CACTCAAACC CCTGGGAACT AACAAGAAAG AAAAAGCGAT AACATTTTAA GTACAATATA CCTCCCCCGT TTCAAAAGTC CCACAACAAA

       100         110         120         130         140         150         160         170         180
AGAATGGGAA GATGTCCCTT GTATCACCAT GGACCCTCAT GATAATTTTG TTTCTTTCAC TTTCTACTCT GTTGACAACC ATTGTCTCCT
TCTTACCCTT CTACAGGGAA CATAGTGGTA CCTGGGAGTA CTATTAAAAC AAAGAAAGTG AAAGATGAGA CAACTGTTGG TAACAGAGGA

       190         200         210         220         230         240         250         260         270
CTTATTTTCT TTTCATTTTC TGTAACTTTT TCGTTAAACT TTAGCTTGCA TTTGTAACGA ATTTTTAAAT TCACTTTTGT TTATTTGTCA
GAATAAAAGA AAAGTAAAAG ACATTGAAAA AGCAATTTGA AATCGAACGT AAACATTGCT TAAAAATTTA AGTGAAAACA AATAAACAGT

       280         290         300         310         320         330         340         350         360
GATTGTAAGT ACTTTCTCTA ATCACTTTTT TTTCAAGGCA ATCAGGGTAT ATTATATTGT ACTTCAGCAC AGTTTTAGAG AACAATTGTT
CTAACATTCA TGAAAGAGAT TAGTGAAAAA AAAGTTCCGT TAGTCCCATA TAATATAACA TGAAGTCGTG TCAAAATCTC TTGTTAACAA

       370         380         390         400         410         420         430         440         450
ATAATTAAAT GATAAGGTAG AATATTTCTG CATATAAATT CTGGCTGGCG TGGAAATATT CTTATTGGTA GAAACAACTA CACCCTGGTC
TATTAATTTA CTATTCCATC TTATAAAGAC GTATATTTAA GACCGACCGC ACCTTTATAA GAATAACCAT CTTTGTTGAT GTGGGACCAG

       460         470         480         490         500         510         520         530         540
ATCATCCTGC CTTTCTCTTT ATGGTTACAA TGATATACAC TGTTTGAGAT GAGGATAAAA TACTCTGAGT CCAAACCGGG CCCCTCTGCT
TAGTAGGACG GAAAGAGAAA TACCAATGTT ACTATATGTG ACAAACTCTA CTCCTATTTT ATGAGACTCA GGTTTGGCCC GGGGAGACGA

       550         560         570
AACCATGTTC ATGCCTTCTT CTTTTTCCTA CAG
TTGGTACAAG TACGGAAGAA GAAAAAGGAT GTC
```

Fig. S14. The second intron of rabbit beta hemoglobin (HBB2; GeneID: 100009084) gene (coordinates 31411:31983 of file gb|M18818.1|RABBGLOB). Restriction sites used for cloning are underlined. *MfeI* (starting at nt 353), *PleI* (starting at nt 517), *PspOMI* (starting at nt 528). The sequence where the Lariat forms and the 3' splice region are shown in bold, whereas the branchpoint is underlined (31).

```
 NheI   10  SnaBI  20         30         40         50         60         70         80         90
gctagcaaag tacgtaCCTC AGGTGCAGGC TGCCTATCAG AAGGTGGTGG CTGGTGTGGC CAATGCCCTG GCTCACAAAT ACCACTGAGA
cgatcgtttc atgcatGGAG TCCACGTCCG ACGGATAGTC TTCCACCACC GACCACACCG GTTACGGGAC CGAGTGTTTA TGGTGACTCT

       100        110        120        130        140        150        160        170        180
TCTTTTTCCC TCTGCCAAAA ATTATGGGGA CATCATGAAG CCCCTTGAGC ATCTGACTTC TGGCTAATAA AGGAAATTTA TTTTCATTGC
AGAAAAAGGG AGACGGTTTT TAATACCCCT GTAGTACTTC GGGGAACTCG TAGACTGAAG ACCGATTATT TCCTTTAAAT AAAAGTAACG

       190        200        210        220        230        240        250        260        270
AATAGTGTGT TGGAATTTTT TGTGTCTCTC ACTCGGAAGG ACATATGGGA GGGCAAATCA TTTAAAACAT CAGAATGAGT ATTTGGTTTA
TTATCACACA ACCTTAAAAA ACACAGAGAG TGAGCCTTCC TGTATACCCT CCCGTTTAGT AAATTTTGTA GTCTTACTCA TAAACCAAAT

       280        290        300        310        320        330        340        350        360
GAGTTTGGCA ACATATGCCC ATATGCTGGC TGCCATGAAC AAAGGTTGGC TATAAAGAGG TCATCAGTAT ATGAAACAGC CCCCTGCTGT
CTCAAACCGT TGTATACGGG TATACGACCG ACGGTACTTG TTTCCAACCG ATATTTCTCC AGTAGTCATA TACTTTGTCG GGGGACGACA

       370        380        390        400        410        420        430        440        450
CCATTCCTTA TTCCATAGAA AAGCCTTGAC TTGAGGTTAG ATTTTTTTTA TATTTTGTTT TGTGTTATTT TTTTCTTTAA CATCCCTAAA
GGTAAGGAAT AAGGTATCTT TTCGGAACTG AACTCCAATC TAAAAAAAAT ATAAAACAAA ACACAATAAA AAAAGAAATT GTAGGGATTT

       460        470        480        490        500        510        520        530        540
ATTTTCCTTA CATGTTTTAC TAGCCAGATT TTTCCTCCTC TCCTGACTAC TCCCAGTCAT AGCTGTCCCT CTTCTCTTAT GGAGATCcct
TAAAAGGAAT GTACAAAATG ATCGGTCTAA AAAGGAGGAG AGGACTGATG AGGGTCAGTA TCGACAGGGA GAAGAGAATA CCTCTAGgga

XhoI
cgag
gctc
```

Fig. S15. A variant of the region of the rabbit beta hemoglobin (HBB2) gene (coordinates 32041:32560 of file gb|M18818.1|RABBGLOB) commonly utilized to promote polyadenylation in mammalian expression vectors. Synthetic regions are in lower case; restriction endonuclease recognition sites *NheI*, *SnaBI*, and *XhoI* sites are italicized and underlined. The sequence in bold denotes part of the HBB2 gene that encodes for the C-terminal end of HBB2 protein; since this sequence is placed downstream of *eGFP*'s ORF and past *eGPF*'s stop codon, it is not incorporated into eGFP. The consensus AAUAAA and G/U box sequence elements that comprise the polyadenylation signal are underlined (32).

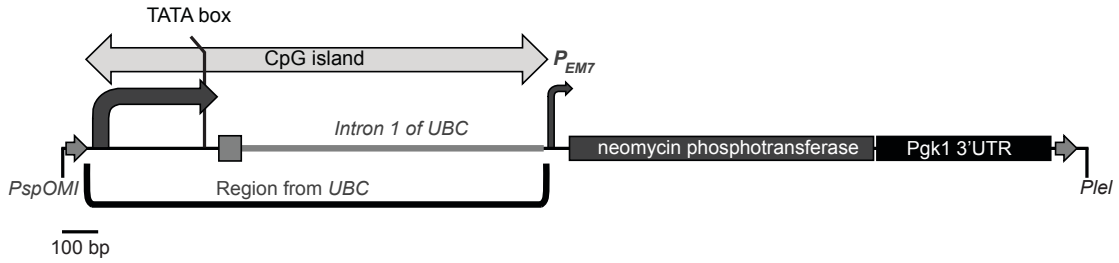


Fig. S16. A UBC/EM7 promoter-driven *neo* mini-gene for expression of Neomycin Phoshotransferase in mammalian cells and *E. coli*. Curved dark gray arrow denotes the human *UBC* promoter, followed by the TATA box, exon 1 (gray box), and intron 1 (thick gray line). This region is contained within a CpG island (light gray bidirectional arrow). An EM7 promoter (P_{EM7}, thin arrow) was engineered immediately downstream of UBC's intron 1 and immediately upstream of the initiating Methionine of *neomycin phosphotransferase* (*npt*, dark gray box) in order to drive expression of *npt* in *E. coli*. The 3'UTR of mouse *Pgk1* (*Pgk1 3'U*, black box) was cloned downstream of *npt* to provide a polyadenylation signal. *FRT* sites (gray arrows) flank *neo* in order to enable its excision using FLP. *PleI* and *PspOMI* restriction sites were used for cloning into the rabbit HBB2 intron 2 (see Fig. 1*A*, and Fig. S1). Scale bar = 100 bp.

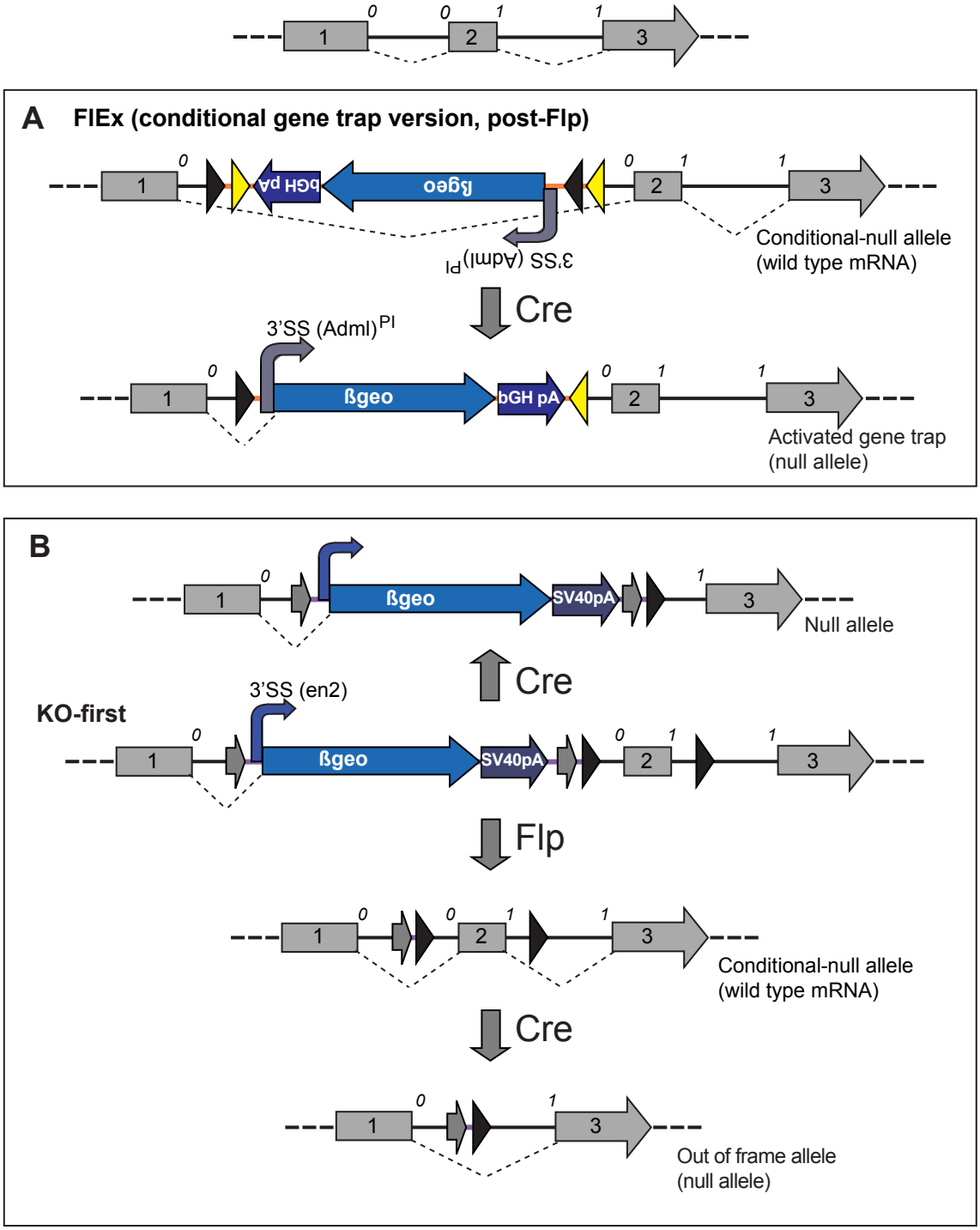


Fig. S17. Comparison of FlEx and KO-first allele design. A hypothetical protein-coding gene amenable to design by KO-first and comprised of three protein-coding exons is depicted. (*A*) FlEx functions as a conditional gene trap. The conditional version of a FlEx conditional gene trap allele is shown, already inverted into the

conditional position (33). Since the 3'SS (Adml)-ßgeo-bGHpA gene trap element is in the antisense strand of the modified gene, the expression of the gene is unaffected. Cre-mediated re-inversion of FlEx's gene trap element places it back into the sense strand, and theoretically results in trapping of the modified gene's message, potentially resulting in a null allele. The *FRT* and *FRT3* sites required for and utilized to generate the conditional version of FlEx are not shown. The 3'SS from the adenovirus major late mRNA intron (4) is depicted as curved gray arrow, whereas the bGH polyA is depicted as dark blue arrow. (*B*) KO-first combines a null with reporter or a conditional-null (without reporter) depending on post-targeting treatment with Flp or Cre. To generate a KO-first allele, a 3'SS (en2)-ßgeo-SV40pA gene trap element (22) is targeted into intron 1, while simultaneously exon 2 is floxed. In this example, exon 2 functions as the critical exon as its excision will result in an out of frame splicing event between exon 1 (which ends in phase 0) and exon 3 (which starts in phase 1); this should result in nonsense-mediated decay (34) of the aberrant mRNA, thereby giving rise to a functional null allele. The initial KO-first allele is converted into a null allele post-targeting, by deleting the critical exon (in this example, exon 2) using Cre. It can also be converted into a conditional-null allele via removal of the gene trap element using Flp. The resulting conditional-null allele relies on deletion of the critical exon to generate a null allele. The 3'SS from *engrailed 2* (en2; region from Chromosome 5, coordinates 28,494,951 to 28,496,659 in Build 58) is depicted as a blue curved arrow, whereas the polyadenylation region from SV40 (SV40pA) (23) is depicted as blue-gray arrow. FRT sites are depicted as short gray arrows. Shared figurative elements and symbols: Exons are shown in gray. The numbers above the exons indicate their coding phase. ßgeo is depicted as a blue arrow. *loxP*, black arrowhead; *lox511*, yellow arrowhead. Dotted line indicates splicing.

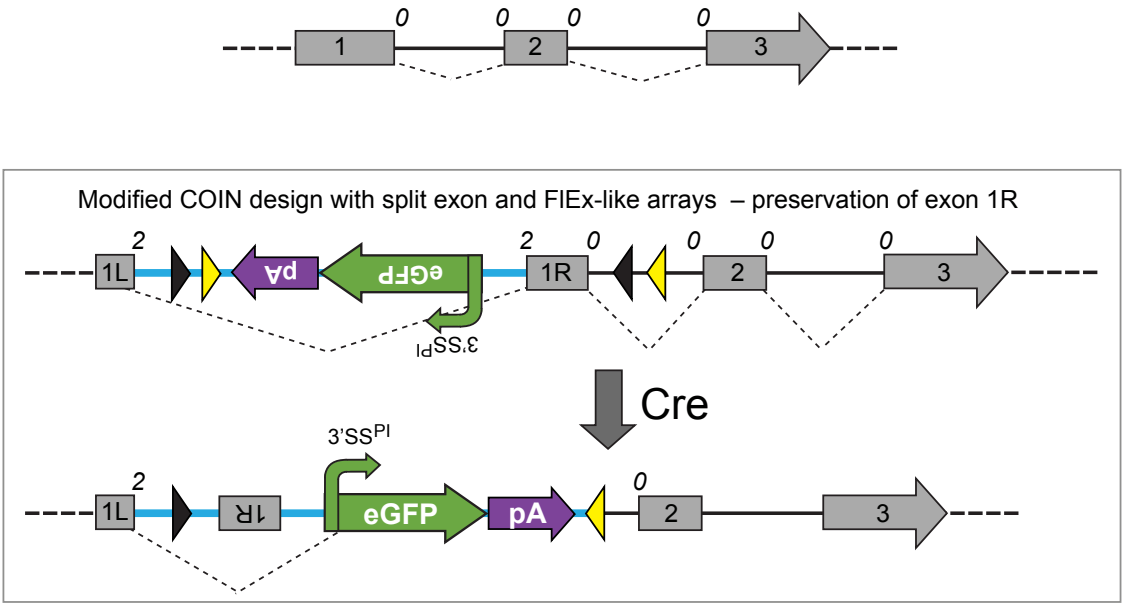


Fig. S18. Converting a non-critical exon to critical using a FlEx-COIN hybrid design: modified COIN design with split exon and FlEx-like arrays – preservation of exon 1R. A hypothetical protein-coding gene comprised of three exons is depicted, where the exons are in the same phase (*phase 0*). Exons are shown blue boxes, splicing in denoted by dotted blue lines and the starting and end phase of each exon is denoted by a number above the corresponding spot on each exon. Exon 1 is rendered "critical" as above, but it is placed between the two FlEx-like arrays. Recombination by Cre results in inversion of the COIN module into the sense strand, and exon 1R into the antisense strand, thereby removing the latter from being incorporated into the mature message. The functional outcome on target gene expression is identical to that in Figure 6*B*. Yellow and purple triangles denote *loxP* and *lox2372* sites respectively, in a FlEx-like array configuration. All other elements are as described in Figure 1.

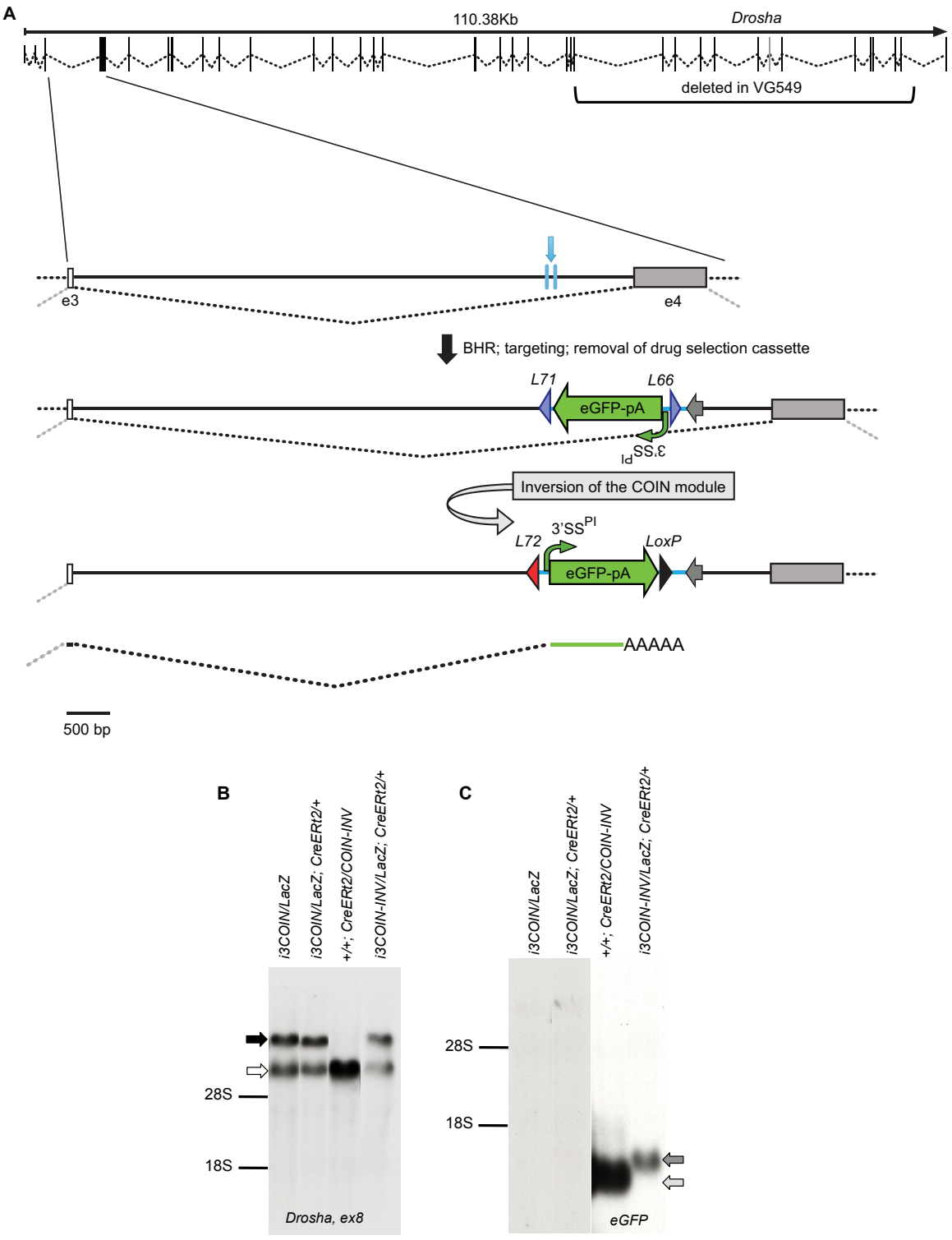


Fig. S19. The intronic COIN of *Drosha* is a hypomorphic, knock-down allele, after inversion of the COIN module. *(A)* Schematic of *Drosha*^{i3COIN} allele. The exon-intron

region of *Drosha* (splice variant 001, transcript ID ENSMUST00000090292) was adapted from Ensembl.org. Light blue vertical arrow indicates the point of insertion of the COIN module within intron 3. Prior to inversion, the *Drosha*^{i3COIN} allele generates a normal message as the COIN intron is spliced out. After inversion, the COIN module should become the last functional exon of the modified gene, abrogating expression of the downstream exons, and resulting in a functional null allele incorporating eGFP. The region replaced in the *Drosha*^{LacZ} allele VG549 is marked by brackets. Naming conventions, abbreviations, and markings for different elements are as noted in Figure 2. Scale bar, 500 bp. (*B*), (*C*) Inversion of the COIN module results in reduction of *Drosha* expression, but both the truncated (*Drosha*$^{i3COIN\text{-}INV}$ mRNA) and the full length mRNA are produced. Northern analysis of Drosha using exon 8 (ENSMUSE000001219097) as a probe reveals reduced expression of Drosha mRNA (white arrow, panel B) in *Drosha*$^{i3COIN\text{-}INV/LacZ}$ ES cells as well as the presence of a hybrid/fusion message encoding exons 1 to 21 of *Drosha* fused to *LacZ* (black arrow, panel *B*). Probing with eGFP reveals the presence of a message comprised of *Drosha*'s exons 1 to 3 fused to eGFP (dark gray arrow, panel *C*). eGFP expressed from the inverted COIN allele of *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)}$ (Supplementary Appendix, Fig. S10) is shown as a positive control for eGFP expression (*C*). The positions of 18S and 28S rRNAs are marked. Genotype key: *i3COIN/LacZ* = *Drosha*$^{i3COIN/LacZ}$, *i3COIN/LacZ* = *Drosha*$^{i3COIN/LacZ}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$, *+/+; CreERt2/COIN-INV* = *Drosha*$^{+/+}$; *Gt(ROSA26)Sor*$^{i1COIN(dsRED2>eGFP)\text{-}INV}$, *i3COIN-INV/LacZ; CreERt2/+* = *Drosha*$^{i3COIN\text{-}INV/LacZ}$; *Gt(ROSA)26Sor*$^{CreERt2/+}$.

Table S1. *Il2rg*$^{ex1COIN\text{-}INV}$*/Y* mice phenocopy *Il2rg*-null mice.

Phenotypes reported for *Il2rg* knockout mice	**Concurrence**
Histologically defined lesions:	nt
Small, hypoplastic thymus (reduced cortex & medula size)	✓
Lymph nodes: • Apparent absence of peripheral and paracortical lymph nodes • Small mesenteric lymph node lacking germinal centers • Mild diffuse proliferation of spindle cells	✓
Gut: absence of lymphoid tissue in small & large intestine	nt
Spleen: • Modest reduction in the size of the spleen • Severe lymphocyte hypoplasia in the white pulp • Absence of B cell and marginal zones	nt
T cell compartment:	
Reduction in T cell number	✓
Decrease in % of $CD4^-$ $CD8^+$ ($CD8^+$ $T_{Cytotoxic}$) cells	✓
Accompanying increase in % of $CD4^+$ $CD8^-$ ($CD4^+$ T_{Helper}) cells	✓
Increase in % of T cells expressing high levels of CD3ε, TCRß, CD5	nt
B cell compartment:	
Marked decrease in B cell number	✓
Near complete absence of $B220^+$ IgM^+ cells in bone marrow	✓
Splenic $B220^+$ IgM^+ cells severely reduced	✓
Splenic $B220^+$ δ^+ cells severely reduced	nt
Normal IgM levels, but all other Ig classes diminished	nt
NK cell compartment:	
Severe decrease in number of NK cells in the spleen	✓

Summary of phenotypes reported for *Il2rg*-null mice (35-38) and comparison with *Il2rg*$^{ex1COIN\text{-}INV}$*/Y* mice. "✓" indicates matching phenotype; nt = not tested. *Il2rg*^{ex1COIN}*/Y* and *Il2rg*$^{ex1COIN\text{-}INV}$*/Y* mice were similar in size and appearance to wild type littermates.

Table S2. *Dll4*$^{i3COIN\text{-}INV/+}$ mice phenocopy *Dll4*$^{LacZ/+}$

Phenotypic tally of *Dll4*$^{i3COIN/+}$ *x Nanog-Cre/+* cross:

Dll4$^{i3COIN/+}$ *Nanog-Cre*		*Dll4*$^{i3COIN/+}$		*Dll4*$^{+/+}$ *Nanog-Cre*		*Dll4*$^{+/+}$	
Affected	Non-affected	Affected	Non-affected	Affected	Non-affected	Affected	Non-affected
11	none	none	11	none	8	none	14

Dll4$^{i3COIN/+}$ mice were mated with a *Nanog-Cre* transgenic line and embryos were collected at E10.5. Embryos were photographed and then genotyped.

Embryo IDs belonging to each genenotypic and phenotypic category:

Dll4$^{i3COIN/+}$; *Nanog-Cre*		*Dll4*$^{i3COIN/+}$		*Dll4*$^{+/+}$; *Nanog-Cre*		*Dll4*$^{+/+}$	
Affected	Non-affected	Affected	Non-affected	Affected	Non-affected	Affected	Non-affected
390619			390620		**392260**		390622
390621			390623		**392261**		391916
390624			391915		392265		391918
390625			391917		392267		391919
391914			391921		394128		391920
392259			392262		394133		391922
392263			392266		394140		392268
392264			392269		394141		394129
394126			**394127**				394130
394142			394131				394131
394143			394136				394134
							394135
							394138
							394139

Embryo IDs noted in bold type indicate that the particular embryo is depicted in Figure 4*B*. Specifically, embryo with ID **392260** is shown on panel **i**, embryo with ID **392261** on panel **v**, embryo with ID **394126** on panels **iii** and **iv**, and embryo with ID **394127** on panels **ii** and **vi**. *Dll4*$^{i3COIN\text{-}INV/+}$ embryos phenocopy correspondingly staged *Dll4*$^{LacZ/+}$ embryos.

Table S3. The *Drosha*$^{i3COIN\text{-}INV}$ allele does not phenocopy the null allele, *Drosha*^{LacZ}.

	***Drosha*$^{LacZ/+}$ x *Drosha*$^{LacZ/+}$ F2 cross**		
	Number of mice per genotypic class		
genotype	*Drosha*$^{+/LacZ}$	*Drosha*$^{LacZ/LacZ}$	*Drosha*$^{+/+}$
expected	111.5	55.75	55.75
observed	151	0	72

The number of mice observed for each genotypic class for a *Drosha*$^{LacZ/+}$ x *Drosha*$^{LacZ/+}$ F2 cross. The parent mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are statistically significant (χ^2 = 74.480; p > 0.0001), driven by a complete absence of the *Drosha*$^{LacZ/LacZ}$ genotypic class. This indicates the deletion-based null allele, *Drosha*^{LacZ}, results in embryonic lethality upon homozygosis, and is consistent with the phenotype that would be expected given that knockouts of *Dgcr8* — the other obligate component of the microprocessor — are also embryonic-lethal (39).

	***Drosha*$^{i3COIN/+}$ x *Drosha*$^{LacZ/+}$ F2 cross**			
	Number of mice per genotypic class			
genotype	*Drosha*$^{i3COIN/+}$	*Drosha*$^{i3COIN/LacZ}$	*Drosha*$^{LacZ/+}$	*Drosha*$^{+/+}$
expected	12.75	12.75	12.75	12.75
observed	15	11	14	11

The number of mice observed for each genotypic class for a *Drosha*$^{i3COIN/+}$ x *Drosha*$^{LacZ/+}$ F2 cross. The parent mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 1.000; p = 0.8013), indicating that the presence of the i3COIN allele of *Drosha* is inherited in normal Mendelian fashion.

	***Drosha*$^{i3COIN\text{-}INV/+}$ x *Drosha*$^{LacZ/+}$ F2 cross**

	Number of mice per genotypic class			
genotype	*Drosha*$^{i3COIN\text{-}INV/+}$	*Drosha*$^{i3COIN\text{-}INV/LacZ}$	*Drosha*$^{LacZ/+}$	*Drosha*$^{+/+}$
expected	24.5	24.5	24.5	24.5
observed	31	7	40	20

The number of mice observed for each genotypic class for a *Drosha*$^{i3COIN\text{-}INV/+}$ x *Drosha*$^{LacZ/+}$ F2 cross. The parent mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are statistically significant (χ^2 = 24.857; p > 0.0001), driven by a reduction in the number of *Drosha*$^{i3COIN\text{-}INV/LacZ}$ mice. However, given the fact that with the deletion-based null allele *Drosha*^{LacZ} no homozygous progeny is born, these results indicate that the i3COIN-INV allele of *Drosha* is hypomorphic rather than correspond to a complete null.

Table S4. The *Drosha*$^{ex4COIN\text{-}INV/LacZ}$ mice are embryonic-lethal

***Drosha*$^{ex4COIN/+}$ x *Drosha*$^{LacZ/+}$ F2 cross**

	Number of mice per genotypic class			
genotype	*Drosha*$^{ex4COIN/+}$	*Drosha*$^{ex4COIN/LacZ}$	*Drosha*$^{LacZ/+}$	*Drosha*$^{+/+}$
expected	41.25	41.25	41.25	41.25
observed	43	39	39	44

The number of mice observed for each genotypic class for a *Drosha*$^{ex4COIN/+}$ x *Drosha*$^{LacZ/+}$ F2 cross. The parent mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 0.503; p = 0.9182), indicating that the presence of the ex4COIN allele of *Drosha* is inherited in normal Mendelian fashion.

***Drosha*$^{ex4COIN/+}$ x *Drosha*$^{LacZ/+}$; *Nanog-Cre* F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Drosha$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre/+*	11.375	0
Drosha$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre/+*	11.375	10
Drosha$^{+/LacZ}$; *Nanog-Cre/+*	11.375	13
Drosha$^{+/+}$; *Nanog-Cre/+*	11.375	5
Drosha$^{ex4COIN/LacZ}$; *+/+*	11.375	14
Drosha$^{ex4COIN/+}$; *+/+*	11.375	17
Drosha$^{+/LacZ}$; *+/+*	11.375	20
Drosha$^{+/+}$; *+/+*	11.375	12

The number of mice observed for each genotypic class for a *Drosha*$^{ex4COIN/+}$ x *Drosha*$^{LacZ/+}$*Nanog-Cre* F2 cross. The *Drosha*$^{LacZ/+}$*Nanog-Cre* mice were in an average genetic background of 79.69% C57BL/6NTac 20.31% 129S6/SvEvTac. The *Drosha*$^{ex4COIN/+}$ mice were in the N2 generation with an average genetic background

of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The difference in the number of mice observed versus expected is statistically significant (χ^2 = 25.308; p = 0.0007). This difference is primarily driven by the absence of the *Drosha*$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre* genotypic class, which is consistent with the *Drosha*$^{ex4COIN\text{-}INV}$ allele being a functional null allele.

Drosha$^{LacZ/ex4COIN}$ x _Drosha_$^{ex4COIN\text{-}INV/+}$; _Nanog-Cre_ cross

	Number of mice per genotypic class	
	Expected	Observed
Drosha$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre/+*	10.375	0
Drosha$^{ex4COIN\text{-}INV/ex4COIN\text{-}INV}$; *Nanog-Cre/+*	10.375	0
Drosha$^{+/LacZ}$; *Nanog-Cre/+*	10.375	14
Drosha$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre/+*	10.375	12
Drosha$^{ex4COIN\text{-}INV/LacZ}$; *+/+*	10.375	0
Drosha$^{ex4COIN\text{-}INV/ex4COIN}$; *+/+*	10.375	15
Drosha$^{ex4COIN/+}$; *+/+*	10.375	26
Drosha$^{+/LacZ}$; *+/+*	10.375	16

The number of mice observed for each genotypic class for a _Drosha_$^{LacZ/e4COIN}$ x _Drosha_$^{ex4COIN\text{-}INV/+}$; _Nanog-Cre_ F2 cross. The *Drosha*$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre* mice were in an average genetic background of 79.69% C57BL/6NTac 20.31% 129S6/SvEvTac. The *Drosha*$^{LacZ/ex4COIN}$ mice were in the N2 generation with an average genetic background of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The difference in the number of mice observed versus expected is statistically significant (χ^2 = 61.289; p < 0.0001). This difference is primarily driven by the absence of the *Drosha*$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre*, *Drosha*$^{ex4COIN\text{-}INV/ex4COIN\text{-}INV}$; *Nanog-Cre*, and *Drosha*$^{ex4COIN\text{-}INV/ex4COIN\text{-}INV}$ genotypic classes. This is consistent with the *Drosha*$^{ex4COIN\text{-}INV}$ allele being a functional null allele.

***Drosha*$^{LacZ/+}$ x *Drosha*$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre* F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Drosha$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre/+*	22.5	0
Drosha$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre/+*	22.5	33
Drosha$^{+/LacZ}$; *Nanog-Cre/+*	22.5	26
Drosha$^{+/+}$; *Nanog-Cre/+*	22.5	35
Drosha$^{ex4COIN\text{-}INV/LacZ}$; *+/+*	22.5	0
Drosha$^{ex4COIN\text{-}INV/+}$; *+/+*	22.5	24
Drosha$^{+/LacZ}$; *+/+*	22.5	40
Drosha$^{+/+}$; *+/+*	22.5	22

The number of mice observed for each genotypic class for a *Drosha*$^{LacZ/+}$ x *Drosha*$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre* F2 cross. The *Drosha*$^{LacZ/+}$ mice were in an average genetic background of 79.69% C57BL/6NTac 20.31% 129S6/SvEvTac. The *Drosha*$^{ex4COIN\text{-}INV/+}$; *Nanog-Cre* mice were in the N2 generation with an average genetic background of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The difference in the number of mice observed versus expected is statistically significant (χ^2 = 71.111; p = 0.0001). In concordance with observations made above, this difference is primarily driven by the absence of the *Drosha*$^{ex4COIN\text{-}INV/LacZ}$; *Nanog-Cre* and *Drosha*$^{ex4COIN\text{-}INV/LacZ}$ genotypic classes, which is also consistent with the *Drosha*$^{ex4COIN\text{-}INV}$ allele being a functional null allele.

Table S5. At the level of inheritance, the *Dicer1*^{i6COIN} allele is wild type whereas the *Dicer1*$^{i6COIN\text{-}INV}$ allele mirrors the null allele, displaying embryonic lethality.

***Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$ F2 cross**

	Number of mice per genotypic class			
genotype	*Dicer1*$^{i6COIN/+}$	*Dicer1*$^{i6COIN/LacZ}$	*Dicer1*$^{LacZ/+}$	*Dicer1*$^{+/+}$
expected	31	31	31	31
observed	34	22	30	38

The number of mice observed for each genotypic class for a *Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$ F2 cross. The parent mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. Although there is trend in the *Dicer1*$^{i6COIN/LacZ}$ mice being under-represented, the differences in the number of mice observed versus expected for each genotypic class are not statistically significant ($\chi^2 = 4.516$; $p = 0.21085612$), indicating that the ***Dicer1*^{i6COIN}** allele is inherited in normal Mendelian fashion.

***Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$; *Nanog-Cre* F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Dicer1$^{i6COIN\text{-}INV/LacZ}$ *Nanog-Cre/+*	9.15	0
Dicer1$^{i6COIN\text{-}INV/+}$ *Nanog-Cre/+*	9.15	10
Dicer1$^{+/LacZ}$ *Nanog-Cre/+*	9.15	5
Dicer1$^{+/+}$ *Nanog-Cre/+*	9.15	10
Dicer1$^{i6COIN/LacZ}$ *+/+*	9.15	10
Dicer1$^{i6COIN/+}$ *+/+*	9.15	8
Dicer1$^{+/LacZ}$ *+/+*	9.15	16
Dicer1$^{+/+}$ *+/+*	9.15	5

The number of mice observed for each genotypic class for a *Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$*Nanog-Cre* F2 cross. The *Dicer1*$^{LacZ/+}$; *Nanog-Cre* mice were in an average genetic background of 79.69% C57BL/6NTac 20.31% 129S6/SvEvTac. The *Dicer1*$^{I6COIN/+}$ mice were in the N2 generation with an average genetic background of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The difference in the number of mice

observed versus expected is statistically significant (χ^2 = 18.424; p = 0.01019647). The difference is driven by the absence of the *Dicer1*$^{i6COIN\text{-}INV/LacZ}$ genotypic class.

***Dicer1*$^{i6COIN\text{-}INV/+}$; *Nanog-Cre* x *Dicer1*$^{LacZ/+}$ F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Dicer1$^{i6COIN\text{-}INV/LacZ}$ *Nanog-Cre/+*	10.33	0
Dicer1$^{i6COIN\text{-}INV/+}$ *Nanog-Cre/+*	10.33	11
Dicer1$^{+/LacZ}$ *Nanog-Cre/+*	10.33	14
Dicer1$^{+/+}$ *Nanog-Cre/+*	10.33	9
Dicer1$^{i6COIN\text{-}INV/LacZ}$ *+/+*	10.33	0
Dicer1$^{i6COIN\text{-}INV/+}$ *+/+*	10.33	3
Dicer1$^{+/LacZ}$ *+/+*	10.33	18
Dicer1$^{+/+}$ *+/+*	10.33	7

The number of mice observed for each genotypic class for a *Dicer1*$^{i6COIN\text{-}INV/+}$ *Nanog-Cre* x *Dicer1*$^{LacZ/+}$ F2 cross. The *Dicer1*$^{LacZ/+}$ mice were in the F1 generation and in an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The *Dicer1*$^{I6COIN\text{-}INV/+}$; *Nanog-Cre* mice were in the F1 generation with an average genetic background of 84.38% C57BL/6NTac 15.62% 129S6/SvEvTac. The difference in the number of mice observed versus expected is statistically significant (χ^2 = 34.148; p = 0.00001616). As in the example above, the difference is driven by the absence of the *Dicer1*$^{i6COIN\text{-}INV/LacZ}$ genotypic class. These results are in agreement with the embryonic lethal phenotype for *Dicer1*-null mice (24) (Fig. S7). Given that *Dicer1*$^{LacZ/+}$ mice are viable and phenotypically normal, it follows that the *Dicer1*$^{i6COIN\text{-}INV}$ allele is most probably a functional null.

***Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$; *Nanog-Cre* F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Dicer1$^{i6COIN\text{-}INV/LacZ}$ *Nanog-Cre/+*	10	0
Dicer1$^{i6COIN\text{-}INV/+}$ *Nanog-Cre/+*	10	10
Dicer1$^{+/LacZ}$ *Nanog-Cre/+*	10	5
Dicer1$^{+/+}$ *Nanog-Cre/+*	10	16
Dicer1$^{i6COIN/LacZ}$ *+/+*	10	10
Dicer1$^{i6COIN/+}$ *+/+*	10	8
Dicer1$^{+/LacZ}$ *+/+*	10	16
Dicer1$^{+/+}$ *+/+*	10	5

The number of mice observed for each genotypic class for a *Dicer1*$^{i6COIN/+}$ x *Dicer1*$^{LacZ/+}$*Nanog-Cre* F2 cross. The *Dicer1*$^{LacZ/+}$; *Nanog-Cre* mice were in the N4 generation with an average genetic background of 96.88% C57BL/6NTac 3.13% 129S6/SvEvTac. The *Dicer1*$^{i6COIN/+}$ mice were in the N2 generation with an average genetic background of 92.19% C57BL/6NTac 7.81% 129S6/SvEvTac. The difference in the number of mice observed versus expected is statistically significant (χ^2 = 22.6; p = 0.002). The difference is driven by the absence of the *Dicer1*$^{i6COIN\text{-}INV/LacZ}$ genotypic class. These results are in agreement with the embryonic lethal phenotype for *Dicer1*-null mice (24) (Fig. S7). Given that *Dicer1*$^{LacZ/+}$ mice are viable and phenotypically normal, it follows that the *Dicer1*$^{i6COIN\text{-}INV}$ allele is most probably a functional null.

Table S6. The *Gdf11*^{ex1COIN} allele displays Mendelian inheritance.

genotype	*Gdf11*$^{ex1COIN/+}$	*Gdf11*$^{ex1COIN/ex1COIN}$	*Gdf11*$^{+/+}$
expected	23	11.5	11.5
observed	19	18	9

The number of progeny observed for each genotypic class born to a *Gdf11*$^{ex1COIN/+}$ x *Gdf11*$^{ex1COIN/+}$ F2 cross. Mice were genotyped at approximately 1 month of age. The parent mice had an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 4.913; p = 0.0857), indicating that homozygosing the COIN allele has no impact on Mendelian inheritance ratios.

Table S7. *Gpr124*$^{ex1COIN/LacZ}$ mice are phenotypically wild type.

	***Gpr124*$^{ex1COIN/+}$ x *Gpr124*$^{LacZ/+}$ F2 cross**			
	Number of mice per genotypic class			
genotype	*Gpr124*$^{ex1COIN/+}$	*Gpr124*$^{ex1COIN/LacZ}$	*Gpr124*$^{LacZ/+}$	*Gpr124*$^{+/+}$
expected	27.5	27.5	27.5	27.5
observed	28	24	30	28

The number of mice observed for each genotypic class for a *Gpr124*$^{ex1COIN/+}$ x *Gpr124*$^{LacZ/+}$ F2 cross. The *Gpr124*$^{ex1COIN/+}$ mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac, whereas the *Gpr124*$^{LacZ/+}$ mice were in the N6 generation with average genetic background of 99% C57BL/6NTac 1% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 0.691; p = 0.8573), indicating that the *Gpr124*^{ex1COIN} allele is wild type.

Table S8. *Plxnd1*$^{ex2COIN/LacZ}$ mice are phenotypically wild type.

	Plxnd1$^{ex2COIN/+}$ **x** ***Plxnd1***$^{LacZ/+}$ **F2 cross**			
	Number of mice per genotypic class			
genotype	*Plxnd1*$^{ex2COIN/+}$	*Plxnd1*$^{ex2COIN/LacZ}$	*Plxnd1*$^{LacZ/+}$	*Plxnd1*$^{+/+}$
expected	25	25	25	25
observed	23	25	22	30

The number of mice observed for each genotypic class for a *Plxnd1*$^{ex2COIN/+}$ x *Plxnd1*$^{LacZ/+}$ F2 cross. The *Plxnd1*$^{ex2COIN/+}$ mice were in the N3 generation with an average genetic background of 93.5% C57BL/6NTac 6.5% 129S6/SvEvTac, whereas the ***Plxnd1***$^{LacZ/+}$ mice were in the F1 generation with an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant ($\chi^2 = 1.520$; $p = 0.6777$), indicating that the ***Plxnd1***^{ex2COIN} allele is inherited in Mendelian fashion.

Table S9. The *Sost iCOIN* and *iCOIN-INV* alleles have no effect on mouse viability.

	***Sost*$^{iCOIN/+}$ x *Sost*$^{LacZ/+}$ F2 cross**			
	Number of mice per genotypic class			
genotype	*Sost*$^{iCOIN/+}$	*Sost*$^{iCOIN/LacZ}$	*Sost*$^{LacZ/+}$	*Sost*$^{+/+}$
expected	66.25	66.25	66.25	66.25
observed	78	69	55	63

The number of mice observed for each genotypic class for a *Sost*$^{iCOIN/+}$ x *Sost*$^{LacZ/+}$ F2 cross. The parent mice had an average genetic background of 75% C57BL/6NTac 25% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 4.268; p = 0.23394873), indicating that the *Sost*^{iCOIN} allele is inherited in Mendelian fashion.

***Sost*$^{iCOIN/+}$ x *Sost*$^{LacZ/+}$*Nanog-Cre* F2 cross**

	Number of mice per genotypic class	
	Expected	Observed
Sost$^{iINV/LacZ}$ *Nanog-Cre/+*	22.5	21
Sost$^{iINV/+}$ *Nanog-Cre/+*	22.5	21
Sost$^{+/LacZ}$ *Nanog-Cre/+*	22.5	27
Sost$^{+/+}$ *Nanog-Cre/+*	22.5	23
Sost$^{iCOIN/LacZ}$ *+/+*	22.5	23
Sost$^{iCOIN/+}$ *+/+*	22.5	19
Sost$^{+/LacZ}$ *+/+*	22.5	18
Sost$^{+/+}$ *+/+*	22.5	26

The number of mice observed for each genotypic class for a *Sost*$^{iCOIN/+}$ x *Sost*$^{LacZ/+}$*Nanog-Cre* F2 cross. The *Sost*$^{LacZ/+}$*Nanog-Cre* mice were in an average genetic background of 93.75% C57BL/6NTac 6.25% 129S6/SvEvTac. The *Sost*$^{iCOIN/+}$ mice were in the N2 generation with an average genetic background of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The difference in the number of mice observed versus expected is not statistically significant (χ^2 = 3.124; p = 0.87333801). This is consistent with the phenotype observed with Sost-null mice

(Fig. S12) and the Mendelian inheritance pattern observed for the *Sost*$^{LacZ/+}$ F2 cross (see below).

	***Sost*$^{LacZ/+}$ x *Sost*$^{LacZ/+}$ F2 cross**		
	Number of mice per genotypic class		
genotype	*Sost*$^{+/LacZ}$	*Sost*$^{LacZ/LacZ}$	*Sost*$^{+/+}$
expected	76.5	38.25	38.25
observed	78	33	42

The number of mice observed for each genotypic class for a *Sost*$^{LacZ/+}$ x *Sost*$^{LacZ/+}$ F2 cross. The parent mice had a genetic background of 100% C57BL/6NTac 25%. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant ($\chi^2 = 1.118$; $p = 0.571881$), indicating that knocking out Sost does not affect Mendelian inheritance ratios.

Table S10. *Tie1*$^{ex1COIN/LacZ}$ mice are phenotypically wild type whereas *Tie1*$^{ex1COIN\text{-}INV/LacZ}$ are embryonic lethal.

	***Tie1*$^{ex1COIN/+}$ x *Tie1*$^{LacZ/+}$ F2 cross**			
	Number of mice per genotypic class			
genotype	*Tie1*$^{ex1COIN/+}$	*Tie1*$^{ex1COIN/LacZ}$	*Tie1*$^{LacZ/+}$	*Tie1*$^{+/+}$
expected	20.25	20.25	20.25	20.25
observed	17	23	17	24

The number of mice observed for each genotypic class for a *Tie1*$^{ex1COIN/+}$ x *Tie1*$^{LacZ/+}$ F2 cross. The parent mice had an average genetic background of 87.5% C57BL/6NTac 12.5% 129S6/SvEvTac. The differences in the number of mice observed versus expected for each genotypic class are not statistically significant (χ^2 = 2.111; p = 0.54966818), indicating that the ***Tie1*^{ex1COIN}** allele is inherited with normal Mendelian frequency.

	***Tie1*$^{ex1COIN\text{-}INV/+}$ x *Tie1*$^{LacZ/+}$ F2 cross**			
	Number of mice per genotypic class			
genotype	*Tie1*$^{ex1COIN\text{-}INV/+}$	*Tie1*$^{ex1COIN\text{-}INV/LacZ}$	*Tie1*$^{LacZ/+}$	*Tie1*$^{+/+}$
expected	13	13	13	13
observed	12	0	11	16

The number of mice observed for each genotypic class for a *Tie1*$^{ex1COIN\text{-}INV/+}$ x *Tie1*$^{LacZ/+}$ F2 cross. The *Tie1*$^{LacZ/+}$ mice were in the N4 generation with an average genetic background of 96.88% C57BL/6NTac 3.13% 129S6/SvEvTac. The *Tie1*$^{ex1COIN\text{-}INV/+}$ mice were in the N2 generation with an average genetic background of 92.19% C57BL/6NTac 7.81% 129S6/SvEvTac. There is a statistically significant difference in the number of mice observed versus expected for the *Tie1*$^{ex1COIN\text{-}INV/LacZ}$ genotypic class (χ^2 = 14.077; p = 0.0028023). The difference is driven by the complete absence of this genotypic class. This is in agreement with the embryonic lethal phenotype for *Tie1*-null mice (Fig. S13) and also expected based on published observations (40). Given that *Tie1*$^{LacZ/+}$ mice are viable and phenotypically normal, it follows that the *Tie1*$^{ex1COIN\text{-}INV}$ allele is most probably a functional null.

Table S11. 3'SS and polyA elements of 3'SS-eGFP-polyA gene-trap cassettes tested

Plasmid name	3' Splice Region (3'SS)	PolyA element	References
pEGFP-N1	n.a.	n.a.	n.a.
pDsRED2-N1	n.a.	n.a.	n.a.
pDsRed2+rßgl	n.a.	n.a.	n.a.
pDsRed2+[rßgl.(L66)/GFP/L71]	rßgl3'SS	none	(21)
pR2i[rßgl.(L66)/GFP-bGHpA/L71]	rßgl3'SS	bGHpA	(2, 21)
pR2i[rßgl.(L66)/GFP-ßglpA/L71	rßgl3'SS	rßglpA	(3, 21)
pR2i[(L66)Adml_GFP-bGHpAL71]	Adml3'SS	bGHpA	(2, 4, 5)
pR2i[(L66)_Adml_GFP_ßglpA-L71]	Adml3'SS	rßglpA	(3, 4)
pR2i[(L66)-en2SAGFP-L71]	En2SA	none	(23)
pR2i[(L66)-en2SAGFPsv40pA-L71]	En2SA	SV40pA	(23)
pR2i[rßgl.(L71)/(GFP-ßglpA)L66]	rßgl3'SS	rßglpA	(3, 21)

The 3'SS and polyA elements used in the different 3'SS-eGFP-polyA gene-trap cassettes tested as described in Figure S11. Abbreviations: 3'SS denotes the 3' splice sites; rßgl3'SS, 3'SS from rabbit beta hemoglobin (HBB2) intron 2; Adml3'SS, 3'SS from Adenovirus major late transcript intron; En2SA, *EcoRI* to 3' splice site region from *engrailed 2* intron (region from Chromosome 5, coordinates 28,494,951 to 28,496,659 in Build 58); bGHpA, bovine growth hormone polyA; rßglpA, rabbit HBB2 polyA; SV40pA, SV40 polyA.

Table S12. *Gdf11*$^{ex1COIN/ex1COIN}$ mice occasionally display an extra rib.

		# of mice with x # of ribs		
Genotype	n	12 ribs	13 ribs	14 ribs
Gdf11$^{+/+}$	10	1	9	-
Gdf11$^{ex1COIN/ex1COIN}$	19	-	13	6

The thoraxes of four-month old mice were stained with Alizarin Red and Alcian Blue and photographed. Whereas *Gdf11*$^{+/+}$ mice display the normal number of ribs, 6 out of 19 *Gdf11*$^{ex1COIN/ex1COIN}$ mice displayed an extra rib. Although this compares favorably to the fact that 89 out of 90 Gdf11 heterozygous-null mice have an extra thoracic rib (26), it also indicates that the *Gdf11*^{ex1COIN} allele may display a certain degree of hypomorphicity.

Table S13. *Gdf11*$^{ex1COIN-INV/ex1COIN-INV}$ newborn pups have short tails.

Gdf11$^{ex1COIN-INV/ex1COIN-INV}$*; Nanog-Cre*	*Gdf11*$^{ex1COIN-INV/+}$*; Nanog-Cre*	*Gdf11*$^{ex1COIN-INV/ex1COIN}$	*Gdf11*$^{ex1COIN/+}$
12/12 had shortened tail	*14/14 had normal tail*	*8/8 had normal tail*	*16/16 had normal tail*

Newborn pups born to a *Gdf11*$^{ex1COIN/ex1COIN}$ x *Gdf11*$^{ex1COIN-INV/+}$; *Nanog-Cre/+* cross were visually scored for the reported short tail phenotype (26) (Fig. S8). The number of mice that displayed the short tail phenotype over the total number of mice scored is listed per genotypic category. Only *Gdf11*$^{ex1COIN-INV/ex1COIN-INV}$ mice displayed this phenotype – mice belonging to the other genotypic categories had no overt phenotypic changes.

REFERENCES

1. Conway L & Wickens M (1985) A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes. *Proc Natl Acad Sci U S A* 82(12):3949-3953.
2. Woychik RP, Lyons RH, Post L, & Rottman FM (1984) Requirement for the 3' flanking region of the bovine growth hormone gene for accurate polyadenylylation. *Proc Natl Acad Sci U S A* 81(13):3944-3948.
3. Lanoix J & Acheson NH (1988) A rabbit beta-globin polyadenylation signal directs efficient termination of transcription of polyomavirus DNA. *EMBO J* 7(8):2515-2522.
4. Niwa M, Rose SD, & Berget SM (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* 4(9):1552-1559.
5. Soriano P, Montgomery C, Geske R, & Bradley A (1991) Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell* 64(4):693-702.
6. Testa G*, et al.* (2004) A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. *Genesis* 38(3):151-158.
7. Buchman AR, Burnett L, & Berg P (1980) The SV40 nucleotide sequence. Appendix A. *DNA Tumor Viruses: Molecular Biology of Tumor Viruses*, ed Tooze J (Cold Spring Harbor Laboratory, New York), 2nd edition Ed Vol 10B, pp 799-841.
8. Valenzuela DM*, et al.* (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* 21(6):652-659.
9. Poueymirou WT*, et al.* (2007) F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat Biotechnol* 25(1):91-99.
10. Joyner AL ed (1999) *Gene Targeting; a Practical Approach* (Oxford University Press Inc., Oxford, New York), Second edition Ed, p 293.
11. Feil R, Wagner J, Metzger D, & Chambon P (1997) Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochemical and biophysical research communications* 237(3):752-757.
12. Crawford LA*, et al.* (2009) Connective tissue growth factor (CTGF) inactivation leads to defects in islet cell lineage allocation and beta-cell proliferation during embryogenesis. *Mol Endocrinol* 23(3):324-336.
13. Canalis E, Zanotti S, Beamer WG, Economides AN, & Smerdel-Ramoya A (2010) Connective Tissue Growth Factor Is Required for Skeletal Development and Postnatal Skeletal Homeostasis in Male Mice. *Endocrinology*.
14. Guney MA*, et al.* (2011) Connective tissue growth factor acts within both endothelial cells and beta cells to promote proliferation of developing beta

cells. *Proceedings of the National Academy of Sciences of the United States of America* 108(37):15242-15247.
15. Anderson KD*, et al.* (2011) Angiogenic sprouting into neural tissue requires Gpr124, an orphan G protein-coupled receptor. *Proc Natl Acad Sci U S A* 108(7):2807-2812.
16. Burge CB, Tuschl T, & Sharp AS (1999) Splicing of Precursors to mRNAs by the Spliceosomes. *The RNA World*, eds Gesteland RF, Cech TR, & Atkins JF (Cold Spring Harbor Press, New York), 2nd Ed, pp 525-560.
17. Schwenk F, Baron U, & Rajewsky K (1995) A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res* 23(24):5080-5081.
18. Puri MC, Rossant J, Alitalo K, Bernstein A, & Partanen J (1995) The receptor tyrosine kinase TIE is required for integrity and survival of vascular endothelial cells. *EMBO J* 14(23):5884-5891.
19. Lobov IB*, et al.* (2007) Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc Natl Acad Sci U S A* 104(9):3219-3224.
20. Graham FL, Smiley J, Russell WC, & Nairn R (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* 36(1):59-74.
21. Buchman AR & Berg P (1988) Comparison of intron-dependent and intron-independent gene expression. *Mol Cell Biol* 8(10):4395-4405.
22. Friedrich G & Soriano P (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* 5(9):1513-1523.
23. Skarnes WC, Moss JE, Hurtley SM, & Beddington RS (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc Natl Acad Sci U S A* 92(14):6592-6596.
24. Bernstein E*, et al.* (2003) Dicer is essential for mouse development. *Nat Genet* 35(3):215-217.
25. Lufkin T*, et al.* (1992) Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. *Nature* 359(6398):835-841.
26. McPherron AC, Lawler AM, & Lee SJ (1999) Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat Genet* 22(3):260-264.
27. Adra CN, Boer PH, & McBurney MW (1987) Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter. *Gene* 60(1):65-74.
28. Duncan GS*, et al.* (1999) Genetic evidence for functional redundancy of Platelet/Endothelial cell adhesion molecule-1 (PECAM-1): CD31-deficient mice reveal PECAM-1-dependent and PECAM-1-independent functions. *J Immunol* 162(5):3022-3030.
29. Li X*, et al.* (2008) Targeted deletion of the sclerostin gene in mice results in increased bone formation and bone strength. *J Bone Miner Res* 23(6):860-869.

30. Sato TN*, et al.* (1995) Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature* 376(6535):70-74.
31. Zeitlin S & Efstratiadis A (1984) In vivo splicing products of the rabbit beta-globin pre-mRNA. *Cell* 39(3 Pt 2):589-602.
32. Gil A & Proudfoot NJ (1984) A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation. *Nature* 312(5993):473-474.
33. Schnutgen F*, et al.* (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* 102(20):7221-7226.
34. Chang YF, Imam JS, & Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76:51-74.
35. Cao X*, et al.* (1995) Defective lymphoid development in mice lacking expression of the common cytokine receptor gamma chain. *Immunity* 2(3):223-238.
36. DiSanto JP, Muller W, Guy-Grand D, Fischer A, & Rajewsky K (1995) Lymphoid development in mice with a targeted deletion of the interleukin 2 receptor gamma chain. *Proc Natl Acad Sci U S A* 92(2):377-381.
37. Ohbo K*, et al.* (1996) Modulation of hematopoiesis in mice with a truncated mutant of the interleukin-2 receptor gamma chain. *Blood* 87(3):956-967.
38. Ikebe M*, et al.* (1997) Lymphohaematopoietic abnormalities and systemic lymphoproliferative disorder in interleukin-2 receptor gamma chain-deficient mice. *Int J Exp Pathol* 78(3):133-148.
39. Wang Y, Medvid R, Melton C, Jaenisch R, & Blelloch R (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 39(3):380-385.
40. Rodewald HR & Sato TN (1996) Tie1, a receptor tyrosine kinase essential for vascular endothelial cell integrity, is not critical for the development of hematopoietic cells. *Oncogene* 12(2):397-404.