bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Collateral damage: Identification and characterisation of spontaneous mutations causing deafness from a targeted knockout programme

Morag A. Lewis[1,2*], Neil J. Ingham[1,2] , Jing Chen[1,2], Selina Pearson[2], Francesca Di Domenico[1], Sohinder Rekhi[1], Rochelle Allen[1], Matthew Drake[1], Annelore Willaert[3], Victoria Rook[1], Johanna Pass[1,2], Thomas Keane[2], David Adams[2], Abigail S. Tucker[4], Jacqueline K. White[2], Karen P. Steel[1,2]

1. Wolfson Centre for Age-Related Diseases, King's College London, London SE1 1UL

2. Wellcome Sanger Institute, Hinxton, CB10 1SA

3. Research Group of Experimental Oto-rhino-laryngology, Department of Neurosciences, KU Leuven – University of Leuven, Leuven, Belgium

4. Centre for Craniofacial and Regenerative Biology, King's College London, London SE1 9RT*
Corresponding author: Morag.lewis@kcl.ac.uk

Acknowledgements

We are grateful to Cassandra Whelan for assistance with the MYO7A staining on the *Atp2b2*^{Tkh} and *Tbx1*^{ttch} mutants, Seham Ebrahim for additional analysis on the *Atp2b2*^{Tkh} mutants, Elysia James for protein modelling of KLHL18, Maria Lachgar-Ruiz for assistance with genotyping, Samoela Rexhaj for help with the mapping of the *rhme* mutation, Hannah Thompson for initial analysis of the *Tbx1*^{ttch} mutants, Zahra Hance for her work on the *vthr* mutant, Rosalind Lacey and James Bussell for assistance with mouse colony management, and the Mouse Genetics Project for initial phenotyping of all these mutants. Scanning electron microscropy was carried out at the Wellcome Sanger Institute and King's College London Centre for Ultrastructural Imaging.

This work was supported by Wellcome (098051, 100669), Medical Research Council (MC_qA137918; G0300212 to KPS.), the European Commission (EUMODIC contract No. LSHG-CT-2006-037188 to KPS), the BBSRC (BB/M02069X/1 to KPS) and King's College London. AST is funded by the Wellcome Trust (102889/Z/13/Z). TK was supported by the Medical Research Council (MR/L007428/1) and BBSRC (BB/M000281/1). AW was supported by a grant from the Research Foundation Flanders (1700317N). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors declare no competing interests.

Author contributions:
Electrophysiology was carried out by NJI, MAL, SP, VR and JP. The mapping and segregation testing was done by MAL, JC, SP, RA and MD, and MAL aligned and analysed the exome sequencing data, and carried out the resequencing. Gross morphological analyses of the middle and inner ear were performed by MAL, AW, SR and MD, and scanning electron microscopy was carried out and analysed by JC, FDD and MD. MAL, FDD, RA, and AW contributed to the immunohistochemistry and qPCR. AST reconstructed and analysed the microCT scans. TK and DA carried out exome sequencing and quality control. MAL, NJI, JC, AST, JKW and KPS supervised the research. MAL, NJI, JC and KPS wrote the paper. All authors edited and approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Mice carrying targeted mutations are important for investigating gene function and the role of genes in disease, but the process of culturing embryonic stem cells during the making of a targeted allele offers opportunities for spontaneous mutations to arise. Identifying spontaneous mutations relies on the detection of phenotypes segregating independently of targeted alleles, and many phenotypes are easy to miss if not specifically looked for. Here we present data from a large, targeted knockout programme in which mice were analysed through a phenotyping pipeline. Twenty-five lines out of 1311 displayed different deafness phenotypes that did not segregate with the targeted allele. We have identified 8 different mutations causing deafness in 16 of these 25 lines and characterised the resulting phenotypes. Our data show that spontaneous mutations with observable effects on phenotype are a common side effect of intensive breeding programmes, including those underlying targeted mutation programmes.

Introduction

Manipulating embryonic stem (ES) cells to insert or alter DNA to study the effect of targeted genes in the resulting organism is a widely practised technique, and the creation and characterisation of mutant organisms is a key step in exploring gene function. The advent of the CRISPR-Cas9 system has prompted exploration of its potential unwanted off-target mutagenic effects, which are of particular concern for its use as a therapeutic tool. While some studies have reported that off-target effects are "minimal and manageable"[1,2], with no increase in mutation frequency[2,3], others have shown that off-target mutations do occur and even alleles present at a low frequency in a G0 mosaic founder could be transmitted to offspring in mice[4]. The creation of targeted mutant alleles involves manipulation of ES cells, and although the mutation frequency is lower in ES cells than in somatic cells[5], mutations do still occur, particularly when the ES cells have been through multiple passages[6]. Spontaneous mutations may thus arise at any point during the process of making a specific allele, including breeding of the resultant organisms.

The Sanger Institute Mouse Genetics Project was a large-scale programme that generated and screened mice carrying knockdown alleles created by the KOMP (Knock Out Mouse Programme) and EUCOMM (European Conditional Mouse Mutagenesis) programmes[7-9]. Each mouse line established carried a single targeted knockout allele, and animals from each line were put through a wide range of phenotyping tests[9]. One of the tests used was the Auditory Brainstem Response (ABR), a highly sensitive test capable of detecting subtle hearing defects[10,11]. Vestibular defects leading to balance problems such as circling and head-bobbing were also noted. This enabled the detection of auditory and vestibular phenotypes which did not segregate with the targeted allele and are likely to have been caused by spontaneous mutations. Of the 1311 lines tested, including 2218 wildtype mice and 6798 mutants, we found 25 lines with non-segregating phenotypes.

We observed a variety of phenotypes by ABR and behavioural assessment, and isolated eight lines showing early-onset severe progressive hearing loss, later-onset progressive hearing loss, low frequency hearing loss, or complete deafness with vestibular dysfunction. Here we present our characterisation of these lines, in which we have identified multiple mutations, some of which affect hearing and/or balance. The causative mutations identified so far include new genes not previously associated with deafness and new alleles of genes known to be involved in hearing. Two of the latter show a phenotype much reduced in severity compared to other mutant alleles of the same gene. To the best of our knowledge, this is the first in-depth study of spontaneous mutations in an intensive targeted mutagenesis programme.

Results

Identification of spontaneous mutations affecting hearing

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

From our routine ABR screening of mice carrying targeted knockout alleles, we identified twenty-two lines where a hearing impairment phenotype did not segregate with the targeted allele (Fig. 1, Table 1). An additional four lines were discovered because they displayed a vestibular defect, one of which (MFFD) had also been through the ABR screen, making twenty-five lines in total. Line or colony names (MXXX) were arbitrarily assigned to each breeding colony to use as part of the unique identifier of each mouse and refer to mice carrying both mutant and wildtype alleles within each colony. After each mouse with an aberrant phenotype was discovered, we screened closely-related mice from the same colony and were able to obtain eight mutant lines displaying reliable inheritance of the observed phenotype (Table 1). We used a standard positional cloning approach to identify the mutations. We set up backcrosses to identify the critical chromosomal region for each mutation (Supplementary Fig. 1) and carried out whole exome sequencing to identify candidate mutations within each region. We resequenced candidate variants by Sanger sequencing and tested the segregation of the candidate mutation with the phenotype. For two mutations we confirmed causation using a complementation test. We were able to identify the causative mutation in all eight lines (Table 1), one of which, *S1pr2*^{stdf}, has already been described[12]. We then confirmed the presence of one of the mutations in 8 more lines, all displaying the same unusual phenotype. Of the 25 lines with aberrant phenotypes, we have identified the causative mutation in 16 of them.

The *Klhl18*^{lowf} allele (MCBX colony): low frequency progressive hearing loss

Mice homozygous for this mutation displayed the unusual phenotype of low frequency hearing loss (*lowf*) (Fig. 2a). We mapped the mutation to a 5.7Mb region on chromosome 9 (Supplementary Fig. 1), in which we found 3 exonic variants (Supplementary Table 1), two of which proved to be false calls when resequenced using Sanger sequencing. The third variant was a missense variant in the gene *Klhl18*, g.9:110455454C>A, causing an amino acid change of p.(Val55Phe) (ENSMUST00000068025) (Fig. 2b, c). We used Phyre2[13] to create a model of KLHL18 based on two structures, a *Plasmodium falciparum* Kelch protein[14] and the crystal structure of the BTB-BACK domains of human KLHL11[15] (Fig. 2d). The affected residue lies in the BTB domain, which is involved in protein-protein interaction, including homodimerisation[16-18]. The mutant phenylalanine, being much larger than the wildtype valine, could potentially disrupt the BTB domain and thus protein function. We confirmed that this was the causative mutation by complementation testing with mice carrying the *Klhl18*$^{tm1a(KOMP)Wtsi}$ targeted allele[9,11] (Fig. 2e). Middle ear dissection and inner ear clearing showed no gross malformations of the ossicles or the inner ear (Supplementary Figs 2, 3). This phenotype was observed in eleven different lines in total (Fig. 1), so we sequenced affected mice from eight more of these lines (in addition to the MCBX line) and discovered that all the affected mice (n=15) were homozygous for the *Klhl18*^{lowf} allele, while unaffected littermates (n=8) were heterozygous or wildtype. We were not able to obtain DNA samples from affected mice of the remaining two lines (MCVL, METD). Further characterisation of the unusual *Klhl18*^{lowf} phenotype is described in [19].

The *Atp2b2*^{Tkh} allele (MEBJ colony): semidominant progressive hearing loss

The Tikho (*Tkh*, Russian for "quiet") mutation was mapped to a 5.2Mb region on chromosome 6 (Supplementary Fig. 1). We found one exonic SNV in this region (Supplementary Table 1), a missense mutation in the *Atp2b2* gene, g.6:113759212G>C, causing an amino acid change of p.(Arg969Gly) (ENSMUST00000101045) (Fig. 3a, b). Mice homozygous for this allele exhibited rapidly progressive hearing loss, while heterozygotes displayed slower progressive hearing loss, with the high frequencies affected first (Fig. 3c, Supplementary Fig. 4). Homozygotes and heterozygotes displayed normal gait and balance. We sequenced mice from the breeding colony and confirmed the segregation of the allele with the phenotypes observed. No gross malformations of the ossicles or inner ear were observed (Supplementary Figs 2, 3). We used qPCR to test RNA expression and immunohistochemistry to study protein localisation, but found no difference between wildtypes,

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

heterozygotes and homozygotes in either test (Fig. 3d, e), suggesting that the variant does not affect mRNA transcription or protein localisation. We modelled the mutation using a model of the rabbit skeletal muscle Ca^{2+}-ATPase[20], which matches 75% of the residues with 100% confidence. This includes the mutant residue, which lies in the cytoplasmic domain between transmembrane domains 8 and 9. There are ten point mutations resulting in amino acid substitutions reported for *Atp2b2* in the mouse (http://www.informatics.jax.org/allele/summary?markerId=MGI:105368, accessed November 2020), all of which result in deafness (*Deaf11*, *Deaf13*[21], *wri*[22], *tmy*[23], *m1Mae*[24], *Elfin*[25], *dfw*[26], *Obv*[27], *m1Btlr* (https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=1638) and *m2Btlr* (https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=3670)). The three closest are *Obv*, *m1Btlr* and *Deaf11*, all of which lie within 100 amino acids of the *Tkh* residue (Fig. 3f, g).

The *Del(10Map3k5-Map7)2Kcl* allele (MEEK colony): progressive hearing loss and male sterility

The rhyme (*rhme*) allele was observed to cause male sterility; no pregnancies or pups were obtained from twelve different affected males, paired in matings for at least 66 days. Affected females were fully fertile. We mapped the mutation to a 7MB region on chromosome 10 (Supplementary Fig. 1), but there were no exonic variants in the region, only an interchromosomal translocation from chr10 to chr13 predicted by BreakDancer (Supplementary Table 1), which proved to be a false call when we carried out Sanger sequencing across the predicted breakpoint. We then looked for deletions of whole exons, using the Integrative Genomics Viewer (IGV) to view all reads aligned to the region, and found a large candidate deletion which we confirmed by segregation testing in phenotyped mice. The causative mutation is a 36.7kb deletion, g.10:20116294_20153024del, which includes the last 9 exons of *Map3k5* and the first exon of *Map7* (Fig. 4a). Mutations in *Map7* have been associated with male sterility[28,29]. Homozygotes for the *rhme* deletion had raised thresholds at high frequencies at 4 weeks old, and the hearing loss progressed with age (Fig. 4b, Supplementary Fig. 4). No gross malformations of the ossicles or inner ear were observed (Supplementary Figs 2, 3). We investigated the hair cells of affected adults using immunohistochemistry and found the organ of Corti was disrupted towards the basal regions, with loss of outer hair cells and disruption and collapse of the tunnel of Corti (Fig. 4c).

The *Tbx1*^{ttch} allele (MDLY colony): complete deafness with vestibular dysfunction

Mice homozygous for the twitch (*ttch*) allele exhibited circling and head bobbing behaviour and had no response to any stimulus up to 95dB, the maximum sound output of our equipment (Fig. 5a). The ossicles were normal in appearance (Supplementary Fig. 2), but we observed signs of inflammation in the middle ear at postnatal day (P)28 (serous effusion, thickened epithelia and capillary hyperplasia) which were more common in affected mice. The gross morphology of the vestibular region was severely affected, with very thin or absent semicircular canals (Fig. 5g). The inner ears of P4 pups displayed a reduced scala media and a thinner stria vascularis at early postnatal stages, although the hair cells appear to have developed normally (Fig. 5d). The saccule had collapsed, but the utricular lumen remained open (Fig. 5d). At adult stages the scala media was even smaller, with a thin or absent stria vascularis, a collapsed Reissner's membrane, extensive degeneration of the organ of Corti, and spiral ganglion cell loss (Fig. 5e). Both utriculus and sacculus had collapsed, and the hair cells in the utricle appeared disorganised (Fig. 5e). We mapped the mutation to a 3Mb region on chromosome 16 (Supplementary Fig. 1), which contained only 1 exonic SNV (Supplementary Table 1), which segregated with the phenotype in the colony. This was a missense variant in *Tbx1*, g.16:18584128C>T, which results in an amino acid change of p.(Asp212Asn) (ENSMUST00000232335) (Fig. 5b, c). Asp212 is the same amino acid affected by the recently-reported nmf219 mutation[30], although the underlying coding change (g.16:18584127T>C), and the resulting amino acid change (Asp212Gly), both differ (Fig. 5f). We carried out a complementation

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

test with the *Tbx1*^{tm1Bld} allele[31] and found that the inner ears of compound heterozygotes were notably smaller than those of littermates carrying one copy of the *ttch* allele and had malformed semicircular canals (Fig. 5f). This phenotype, which resembled that of *Tbx1*^{ttch} homozygotes, confirmed that the *ttch* allele is the causative mutation.

We also identified two other variants in this line before the non-recombinant region was fully defined; a 27bp inframe deletion in the *Kmt2d* gene which underlies Kabuki syndrome in humans[32] (Supplementary Fig. 5a), and a missense mutation in the gene *Muc13* (Supplementary Fig. 5c, d). These mutations were confirmed to be present in multiple mice from the MDLY colony but did not segregate with the *ttch* phenotype, and follow-up ABR tests found no effect on hearing in homozygotes of either mutation (Supplementary Fig. 5b, e, g). We investigated the expression of MUC13 in the cochlea at P4 and found it was expressed in hair cells, pillar cells and the basal cells of the stria vascularis (Supplementary Fig. 5f).

The *Pcdh15*^{jigl} allele (MEWY colony): complete deafness with vestibular dysfunction

The jiggle (*jigl)* allele was mapped to a 26.9Mb region on chromosome 10 (Supplementary Fig. 1) in which we found no small exonic variants, but one potential interchromosomal translocation within the gene *Eef2* was identified by BreakDancer (Supplementary Table 1). We investigated this using IGV and found that it was based on a small percentage of the reads covering the region, most of which were correctly mapped, suggesting it was a false call. We then examined the entire non-recombinant region and identified a deletion towards the 3' end of *Pcdh15*, which includes up to 6 coding exons depending on the transcript. There are 29 protein-coding isoforms of *Pcdh15* (ensembl.org, accessed July 2021), 22 of which contain the affected exons. The deletion results in the loss of the 3' end of the coding sequence for eleven of those transcripts, and in the loss of internal exons for the remaining eleven (Fig. 6a). We localised the 5' breakpoint to the region between 10:74614441 and 10:74619826, and the 3' breakpoint to the region between 10:74634994 and 10:74635149. We used primers designed to amplify a region within the deletion (Supplementary Table 3) to confirm it was not present in 20 affected animals, with 20 unaffected mice from the MEWY colony as controls. *Pcdh15*^{jigl} homozygotes display circling and head bobbing and have no auditory brainstem response to any stimulus up to 95dB (Fig. 6b). There were no gross malformations of the ossicles or inner ear (Supplementary Figs 2, 3). Scanning electron microscopy of the organ of Corti at P30 showed that affected mice exhibited hair bundle disorganisation that was most marked in the outer hair cells (Fig. 6c). We observed a similar phenotype in the P5 organ of Corti, with disorganisation of the stereocilia within hair cell bundles, a distorted bundle shape overall, and disoriented bundles (Fig. 6c, d). Hair bundles in the P5 vestibular maculae also lacked the typical staircase organisation (Fig. 6d). These hair bundle defects are similar to those described in other *Pcdh15* mutants and reflect the role of PCDH15 as a component of the tip links between adjacent stereocilia[33,34].

The *Del(18Ctxn3-Ccdc192)1Kcl* allele (MFFD colony): complete deafness with vestibular dysfunction

Mice homozygous for the rhythm (*rthm*) allele exhibited complete deafness (Fig 7a) with circling and head bobbing, suggesting vestibular dysfunction. We mapped the mutation to a 3.3Mb region on chromosome 18 (Supplementary Fig. 1) but did not identify any exonic variants in the region (Supplementary Table 1). We used IGV to examine the nonrecombinant region and discovered a 303kb deletion on chromosome 18, g.18:57437258_57740507del, covering eight genes, including two protein-coding genes (*Ctxn3*, *Ccdc192*), four lncRNA genes, one miRNA and one snRNA (Fig. 7c). This deletion segregated with the phenotype within the colony. The ossicles appeared normal (Supplementary Fig. 2) but the lateral semicircular canal was thinner in homozygotes than in heterozygotes (Fig. 7d), and Myo7a staining revealed severe disruption of the cochlear duct (Fig. 7e). This phenotype resembles that of mice mutant for *Slc12a2*[35], which lies 138kbp 3' of the deletion, so

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

we extracted RNA from brain tissue from affected mice and their unaffected littermates at four weeks old, and carried out qPCR to determine whether *Slc12a2* was misregulated. However, there were no significant differences in the levels of *Slc12a2* expression in homozygotes compared to heterozygote littermates (Fig. 7b). Thus, we found no evidence of a position effect of the deletion on *Slc12a2* expression in brain, suggesting it is more likely that a gene within the deletion is responsible for the phenotype.

The *Espn*^{spdz} allele (MHER colony): complete deafness with vestibular dysfunction

Mice homozygous for this mutation displayed circling and head bobbing and had no response to sound up to 95dB (Fig. 8a); the allele was named spindizzy (*spdz*). There were no gross malformations of the ossicles and inner ear (Supplementary Figs 2, 3), and MYO7A staining in adults suggested hair cells were present (Fig. 8b). However, scanning electron microscopy showed that there were no stereocilia bundles visible at P28 in homozygotes (Fig. 8c), and at P4, stereocilia bundles were present but disorganised, with thin stereocilia and ectopic stereocilia rows (Fig. 8d). The mutation mapped to a 3Mb region on chromosome 4 (Supplementary Fig. 1) containing 28 protein-coding genes. There were 13 small variants called by Samtools in this region (Supplementary Table 1), none of which were within coding sequence. BreakDancer detected four intrachromosomal rearrangements (Supplementary Table 1), affecting exons of *Klhl21*, *Nol9*, *Acot7*, and an intron of *Plekhg5*, but all four are based on a very low proportion of the reads in each homozygote. The only known deafness gene in the non-recombinant region is *Espn*, mutations in which result in shorter, thinner stereocilia at birth, followed by degeneration of stereocilia in early adulthood[36-38]. We therefore sequenced *Espn* mRNA from the brains of adult affected mice and their unaffected littermates. No splicing errors were observed for most of the exons; however, we were unable to amplify sequence from exons 15 and 16 in homozygotes (Fig. 8e). Exon 15 (ENSMUSE00001290053) is a 12bp exon located 85bp from exon 14, and separated by 2.35kb from exon 16, which is the final exon (Fig. 8e). We resequenced the genomic region between exons 14 and 16 in three *spdz* homozygotes using Sanger sequencing, and while no variants were observed in intron 14-15 or in exon 15, we were unable to amplify a 431bp region in the intron between exons 15 and 16, between g.4:152122586 and g.4:152123017, in any of the homozygotes. We successfully sequenced this region in two wildtype mice. This failure to amplify suggests that there may be an insertion or other genomic disruption at this location of the chromosome. We extracted DNA from 55 affected mice from the colony, and in all of them this specific sequence failed to amplify, while an adjacent sequence worked. In 56 unaffected mice both sequences were amplified (Fig. 8e,f; primers in Supplementary Table 3).

Three other lines which underwent exome sequencing

In addition to the above seven lines, mice from three other lines displaying nonsegregating phenotypes were sequenced (MAKN, MATH and MBVF colonies, Fig. 1). The mutation underlying the rapidly progressive hearing loss phenotype seen in the MAKN line was a point mutation in the *S1pr2* gene, named stonedeaf (*S1pr2*^{stdf}). Mice homozygous for this allele displayed a rapid reduction in endocochlear potential (EP) between P14 and P56 which correlated with the progression of their hearing loss, while hair cell degeneration followed at a later age[12].

In the case of the mutation in the MATH line we were not able to establish a breeding colony inheriting the phenotype, but we were able to confirm the presence of the *Klhl18*^{lowf} mutation from the exome sequence, which is in accordance with the low frequency hearing loss exhibited by these mice (Fig 1).

The affected mice from the MBVF line displayed variably raised thresholds across all frequencies (Fig 1); we named the allele variable thresholds, *vthr*. However, because of this extreme variability, we were unable to carry out a backcross or maintain the *vthr* phenotype within the colony. We did

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

observe that the thresholds of 7 affected mice, while variable at 14 weeks old, progressed to more severe hearing loss at 6 months old (Supplementary Fig 4c). We collected and examined the middle ears of these affected mice and related unaffected mice at ages over 6 months and did not observe any middle ear defects (n=7 affected, 17 unaffected). Because of this variability in ABR thresholds, we did not restrict the zygosity of identified variants during variant processing (Supplementary Table 1). After quality and impact filtering, we found 221 potential high impact variants, including 15 large structural variants that were predicted to affect 30 known deafness genes between them (Supplementary Table 2).

ES Cell sequencing

The 25 lines with spontaneous mutations were derived from several different parental embryonic stem cell lines (Table 1). We carried out whole exome sequencing on three of these lines (JM8.F6, JM8.N4 and JM8.N19) and used Sanger sequencing of selected regions to check two others (JM8A1.N3 and JM8A3.N1)[39]. For the eight alleles affecting hearing described above (including *S1pr2*^{stdf}), we found that the mutations were not present in the parental ES cell lines (Table 1). The *Klhl18*^{lowf} mutation, the only one seen in multiple lines, was not found in any of the ES cell lines from which we obtained sequence (Table 1).

In order to find out whether any variants seen in the mice could have been derived from the parental ES cell lines, we compared the JM8.F6 and JM8.N4 whole exome sequencing to whole exome sequencing from four descendant mice (two from the MCBX line and two from the MATH line, both of which carried the spontaneous *Klhl18*^{lowf} mutation, Fig 1, Table 1). For this we adapted the variant filtering steps and processed exome sequence data from each mouse independently (Supplementary Table 4a). We compared the high-quality variants identified by the different callers and found that a subset of ES cell variants was indeed found in the mice created from each line (e.g. 8 out of 1105 variants called by SAMtools were found in the JM8.F6 ES cells and in both MCBX mice, Supplementary Fig 6). We chose 21 high quality, high impact variants for confirmation by Sanger sequencing but most variants were not validated in either the ES cells or the mice (Supplementary Table 4b). However, two variants were found in both JM8.F6 and the two MCBX mice; one point mutation identified by SAMtools and one small indel called by Dindel (Supplementary Table 4b). These variants were not seen in the two MATH mice, which shared the same low frequency hearing loss mutation and the spontaneous *Klhl18*^{lowf} allele, nor in the JM8.N4 ES cell sequence.

Potential sources of the spontaneous mutations affecting hearing

As none of the spontaneous mutations we found that affected hearing came from the original parental ES cell lines (prior to genetic manipulation), we examined the pedigrees of the mice originally found to have these phenotypes and the mice used as founders for our eight breeding colonies. For each line, we identified the latest possible point at which the mutant allele could have arisen, assuming that it only occurred once. In five lines, the mutation could have arisen just two generations before it was observed (MEBJ (*Tkh*), MHER (*spdz*), MEEK (*rhme*), MDLY (*ttch*) and MEWY (*jigl*)). In the MAKN (*stdf*) line, the mutation must have arisen no less than three generations before the phenotype was observed (Supplementary Fig 7). However, in the MFFD (*rthm*) line, the mutation must have been present in the chimaeric offspring of the microinjected founder (Supplementary Fig. 7). For the lines bearing the *Klhl18*^{lowf} allele, the most likely source is the wildtype colony used for embryo donors, microinjection and colony expansion. We checked the pedigrees of the mice homozygous for the *Klhl18*^{lowf} allele and confirmed that in each case there was an ancestral mouse from the same wildtype colony which could have passed on the *Klhl18*^{lowf} allele to both the dam and sire of the homozygous mouse. We then constructed pedigrees for these wildtype mice and determined that it is likely that the *Klhl18*^{lowf} allele was present in several of the founders of the C57BL/6N wildtype colony used to expand the mutant lines[9].

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In summary, one mutation was present at the point of microinjection (MFFD, *rthm*) and therefore arose when the ES cells were targeted or during the ES cell processing prior to microinjection, and one is likely to have arisen spontaneously in a wildtype colony (MCBX, *Klhl18*^{lowf}). The remaining six mutations could have occurred either during ES cell targeting and processing, or during breeding of the colony carrying a targeted allele (Fig 9).

Discussion

Here we have described seven mutant alleles affecting hearing which, along with the previously described *S1pr2*^{stdf} mutation[12], arose as spontaneous mutations within a targeted knockout programme (Table 1). In total, we observed 25 lines with hearing impairment which did not segregate with the targeted allele (Fig. 1), and we identified the causative mutation in 16 of them. It's likely that in the cases where we could not establish a breeding colony carrying the phenotype, the related mice we obtained weren't carrying the mutation. However, it's possible that in the lines where only one mouse was found to be affected, the mutation causing hearing loss was a somatic mutation (eg MGKQ, Fig. 1). In one case (MBVF), the phenotype was variable, and could not be reliably maintained. We have identified six novel alleles of known deafness genes (*Klhl18* [10,11], *S1pr2* [12,40-42], *Atp2b2*[21-27,43], *Tbx1*[44-47] *Pcdh15*[48] and *Espn*[49]), and four candidate deafness genes (*Ctxn3*, *Ccdc192*, *Map3k5* and *Map7*). These last lie within the *rthm* and *rhme* deletions in the MFFD and MEEK lines, and more study is required to identify which gene is responsible for the hearing phenotype or whether, in the case of the *rthm* deletion, it is one of the noncoding RNA genes that underlies the deafness and vestibular dysfunction in this line.

Of the known deafness genes, two of our mutations are hypomorphs, *Atp2b2*^{Tkh} (MEBJ) and *Tbx1*^{ttch} (MDLY) (Table 1). In the case of *Tbx1*, mice homozygous for a null allele die at birth due to failure to inflate their lungs, but gross morphological abnormalities are evident as early as E8.5[44]. mRNA dosage experiments have shown that a very small reduction below the mRNA expression level seen in heterozygotes is sufficient to induce full neonatal lethality[50]. One other *Tbx1* hypomorph has been reported, the *Tbx1*^{nmf219} mutant, which carries a missense mutation affecting the same amino acid as the *Tbx1*^{ttch} allele, Asp212[30]. *Tbx1*^{nmf219} and *Tbx1*^{ttch} homozygotes display a similar inner ear phenotype, with hypoplasia of the semicircular canals, a reduced scala media, collapsed Reissner's membrane, reduced or absent stria vascularis, degeneration of the organ of Corti, and spiral ganglion cell loss[30]. TBX1 binds to DNA as a dimer, and Asp212 is located between the two monomers, so amino acid changes could potentially affect the structure of the dimer and its capacity to bind DNA[30] (Fig. 5c). It is interesting to note that while many other mutant alleles of *Tbx1* are homozygous perinatal lethal[31,44,45], including at least one point mutation[51], neither of these Asp212 mutants show reduced viability of homozygotes. *Atp2b2* mutants also display a difference in phenotype depending on the level of activity of PMCA2, the encoded protein[52]. There are 15 reported *Atp2b2* mutants with phenotypes, 10 of which result from a single amino acid change (Fig 3, http://informatics.jax.org, accessed December 2020). Only the *Deaf11* and *Deaf13* missense alleles are like the *Tkh* allele in that they do not result in ataxia in homozygotes[21-27,43,52-54] ((https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=1638) and (https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=3670)). The new hypomorphic alleles reported here may be useful for in-depth studies of the function of these genes.

It is likely that over a thousand genes contribute to the development and function of the ear, making hearing impairment relatively sensitive to background mutation rates[9,11]. From the 9016 mice (2218 wildtype mice and 6798 targeted mutants) screened by ABR in the current study, 55 had a non-segregating phenotype (0.6%). However, seven of the eight mutations causing a hearing phenotype were inherited in a recessive way, so that we had to screen a mouse with two copies to detect it, and only four mice per colony were normally screened by ABR, although in some cases wildtype littermates were included in the control cohort. Thus, the number of new mutations we found

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

causing deafness is likely to be an underestimate, as many colonies harbouring new mutations causing deafness will not have led to a homozygote reaching the screening cohort. Like many phenotypes, hearing impairment in a mouse, when unaccompanied by vestibular dysfunction, is a subtle phenotype and easy to miss if it is not actively investigated. It is highly likely that multiple other mutations which have no impact on hearing nor any gross impact on the appearance, behaviour or viability of the mice exist within all these lines. We found two such mutations in the process of sequencing the MDLY (*ttch*) mice: a missense mutation in *Muc13* and a deletion in the Kabuki syndrome gene *Kmt2d* (also known as *Mll2*), neither of which affected hearing nor had any obvious effect on homozygotes (Supplementary Fig. 5).

When we first observed the non-segregating phenotypes, particularly the widespread *Klhl18*^{lowf} mutation, we suspected that the mutations arose in the ES cells used to make the mice, either before or after insertion of the manipulated allele. Indeed, a previous missense mutation in *Atp2b2* has been reported to have arisen during clonal expansion of targeted ES cells[24], and multiple other examples of off-target mutations arising in targeted ES cells have been described[55,56]. However, when we sequenced the parental ES cell lines, we did not find any of these mutations, and from our pedigree analyses, only one can be unambiguously traced to the founder of the line (MFFD, *rthm*), although since we can only identify the latest possible time of occurrence from the pedigrees, we can't rule out the possibility that the other mutations occurred within the targeted ES cell. In addition, the presence of the *Klhl18*^{lowf} mutation in multiple lines suggests that it arose within the wildtype colony used for expansion of the mutant lines (Fig 9). We therefore suggest that it is likely that at least some of the mutations affecting hearing arose during breeding. Thus, regardless of the method of genome manipulation or how long ago a mutation was made, the potential for spontaneous and off-target mutations to affect a mutant animal being studied must always be borne in mind. In particular, a non-segregating mutation that affects the phenotype under study could result in apparently variable penetrance or expressivity of the phenotype. It's possible that the variable phenotype observed in the MBVF (*vthr*) line (Fig. 1) is the result of more than one spontaneous mutation.

As well as being a reminder to pay attention to unexpected phenotypes, these results highlight the critical importance of proper data collection and retention in any study. The accurate breeding records and careful tracking of the thousands of mice tested in the Mouse Genetics Project were critical to the detection, isolation and identification of these new mutations. Furthermore, confirmation of candidate variants by Sanger sequencing was essential to identify false calls from exome sequence.

Methods

Ethics statement

Mouse studies were carried out in accordance with UK Home Office regulations and the UK Animals (Scientific Procedures) Act of 1986 (ASPA) under UK Home Office licences, and the study was approved by both the Wellcome Trust Sanger Institute and the King's College London Ethical Review Committees. Mice were culled using methods approved under these licences to minimize any possibility of suffering.

Generation of mice

Mice carrying knockout first conditional-ready alleles were observed for gross behavioural abnormalities using a modified SHIRPA test at 9 weeks old, and tested by ABR at 14 weeks old, as described in [9] and [11]. Mice assigned to the phenotyping pipeline could not be withdrawn for investigation of these non-segregating phenotypes, so mice from the same line, as closely related to

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the affected mice as possible, were used to set up colonies to screen for the spontaneous mutation. 25 lines were found to carry an spontaneous mutation involving hearing impairment, and eight new breeding colonies were successfully established with the observed phenotype reliably inherited. All mutations were generated and maintained on the C57BL/6N background. These eight mouse lines will be available via the European Mouse Mutant Archive. Two further mouse lines were used for complementation testing of compound heterozygotes: *Klhl18*$^{tm1a(KOMP)Wtsi}$ [9] and *Tbx1*^{tm1Bld} [31].

Auditory Brainstem Response (ABR)

ABR tests were carried out as previously described[11,57,58]. We used a broadband click stimulus and shaped tonebursts at a range of pure tone frequencies at sound levels from 0-95dB SPL, in 5dB steps. 256 sweeps were carried out per frequency and sound level, and these were averaged to produce the ABR waveform. A stack of response waveforms was used to identify the threshold for each stimulus, the lowest intensity at which a waveform could be distinguished.

Assessment of balance

Mouse behaviour was observed in their home cage for signs of circling or headbobbing. To detect balance dysfunction, we also used the contact righting reflex test. Mice were placed in a large petri dish sized such that the back of the mouse was in contact with the dish lid but not under pressure. The petri dish was inverted, and the mouse was monitored for 30 seconds. If it did not turn itself over within that time, it was marked as affected, while if it turned itself over promptly, it was marked as unaffected. Some mice righted themselves after a delay and were tested a second time. If the results of the righting test were unclear, the mouse was not used for mapping.

Linkage mapping

For each line, affected males were outcrossed to C3HeB/FeJ females. For the lines displaying recessive inheritance, the offspring were backcrossed to affected mice of the same line. For the single line displaying semidominant inheritance (MEBJ, *Atp2b2*^{Tkh}), the outcross offspring were backcrossed to C3HeB/FeJ wildtype mice. Backcross offspring were assessed by ABR or contact righting reflex, then after culling, tissue was collected for DNA extraction. The initial genome scan was carried out using a standard marker panel (Supplementary Table 5), then once the linkage region was established, it was narrowed down using more backcross mice, more markers within the region and, where necessary, strain-specific SNVs (Supplementary Table 5).

Exome sequencing

Two affected mice of each line were selected for exome sequencing. DNA was extracted using phenol and chloroform, and exome sequencing was carried out by the Wellcome Trust Sanger Institute (WTSI), with the exception of the *spdz* sequencing, which was carried out by Novogene (Hong Kong). Genomic DNA (approximately 1 ug) was fragmented to an average size of 150 bp (WTSI) or 180-280bp (Novogene) and subjected to DNA library creation using the Agilent SureSelect Mouse All Exon Kit. Adapter-ligated libraries were amplified and indexed via PCR. Enriched libraries were subjected to 75bp paired-end sequencing (HiSeq 2000; Illumina), with the exception of the *spdz* libraries, which were subjected to 150bp paired-end sequencing (HiSeq 4000; Illumina) following the manufacturer's instructions. Sequences were aligned to GRCm38 using bwa[59], with the exception of the *spdz* fastq files, which were checked and processed using FastQC[60] and Trimmomatic[61] , then aligned to GRCm38 using hisat2[62]. All bam files were improved by local realignment around insertions and deletions discovered in the mouse genomes project[63] using GATK[64]. Bam files were processed using Picard and Samtools[65,66]. Raw data can be downloaded from the European Nucleotide Archive, studies PRJEB2585, PRJEB5221 and PRJEB45713. Individual sample accession numbers are in Supplementary Table 6c.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Variant identification and investigation

Variant calling was carried out using Samtools, Dindel and Pindel[66-69], and we also used BreakDancer to detect any large structural variants with breakpoints within the exome[68] (Supplementary Table 6).Variants were annotated using the Ensembl Variant Effect Predictor[70]. For the ES cells and the MBVF line we could not use mapping cross information so quality and impact filtering was carried out instead as the final step (Supplementary Table 6). IGV[71] was used to check for large deletions in the critical region. Candidate variants were first confirmed by Sanger sequencing, then checked for segregation with the phenotype across the colony as well as all backcross mice (Supplementary Table 3). Protein modelling was carried out using Phyre2[13] to identify a suitable protein model, and Pymol[72] to view the model and highlight variant residues. All Sanger sequencing was carried out by Source Bioscience and analysed using Gap4[73]. Venn diagrams were generated using the tool at http://bioinformatics.psb.ugent.be/webtools/Venn/.

Scanning Electron Microscopy (SEM)

The temporal bones were isolated. The inner ears were dissected out and fixed by 2.5% glutaraldehyde in 0.1M sodium cacodylate buffer with 3 mM calcium chloride at room temperature for 3 hours. Cochleae and the vestibular system were finely dissected in PBS. This was followed by further processing using an osmium-thiocarbohydrazide-osmium (OTOTO) method[74]. The samples were dehydrated in increasing concentrations of ethanol, critical-point dried (Bal-Tec CPD030), mounted and examined under a HITACHI S-4800 or a JEOL JSM 7800F Prime Schottky field emission scanning electron microscope (for *Pcdh*^{jigl} and *Espn*^{spdz} mice respectively). Images of the organ of Corti were taken at roughly 20% intervals along the cochlear duct and the macula of the utricle and saccule and crista of the ampullae were also imaged. Whole images were adjusted in Photoshop to normalise dynamic range across all panels.

Middle ear dissection, inner ear clearing and microCT scanning

After culling, the ear canals were checked for cerumen, then the mouse was decapitated and the bulla, tympanic membrane and middle ear cavity inspected for any abnormalities, including the presence of fluid, inflammation or any other obstructions in the middle ear. Observations were recorded on a standard tick sheet. Ossicles were dissected out and stored in 10% formalin. The inner ear was removed and fixed in Bodian's fixative (75% ethanol, 5% acetic acid, 5% formalin, in water), washed in water and 70% ethanol, then cleared by gentle rotation in 3% KOH for 3 days, changed daily. Samples were then placed in G:E:B (glycerol, 70% ethanol and benzyl alcohol, mixed in a ratio of 2:2:1 by volume) for the final stage of clearing, then stored in G:E (glycerol and 70% ethanol in equal volumes). Images of ossicles and middle ears were taken using a Leica stereomicroscope with a Leica DFC490 camera. To carry out microCT scans, cochleae were immobilized using cotton gauze and scanned with a Scanco microCT 50 to produce 14µm voxel size volumes, using an X-ray tube voltage of 80kVp and a tube current of 80µA. An aluminium filter (0.05mm) was used to adjust the energy distribution of the X-ray source. To ensure scan consistency, a calibration phantom of known geometry (a dense cylinder) was positioned within the field of acquisition for each scan. Test reconstructions on this object were carried out to determine the optimum conditions for reconstruction, ensuring consistency in image quality, and minimising blurring. Reconstruction of the cochlea was performed in Thermo Scientific Amira software.

Immunohistochemistry and trichrome staining

Samples from adult mice were collected, fixed in 10% formalin, decalcified in 0.1M EDTA, embedded in paraffin wax and cut into 8µm sections. Samples from P4 pups were treated similarly, but no decalcification step was needed. For histological analysis, slides were stained using a trichrome stain, containing Alcian blue, Sirius red and Haematoxylin. Immunohistochemistry was carried out using a

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Ventana Discovery machine and reagents according to the manufacturer's instructions (DABMap™ Kit (cat.no 760-124), Haematoxylin (cat.no 760-2021), Bluing reagent (cat.no 760-2037), CC1 (cat.no 950-124), EZPrep (cat.no 950-100), LCS (cat.no 650-010), RiboWash (cat.no 760-105), Reaction Buffer (cat.no 95-300), and RiboCC (cat.no 760-107)). Primary antibodies used were rabbit anti-PMCA2 (Abcam, cat. no: ab3529, diluted 1:500), rabbit anti-MUC13 (Abcam, cat. no: ab124654) and rabbit anti-MYO7A (Proteus, cat. no: PTS-25-6790, diluted 1:100), and the secondary antibody was anti-rabbit (Jackson ImmunoResearch, cat.no 711-065-152, diluted 1:100). Antibodies were diluted in staining solution (10% foetal calf serum, 0.1% Triton, 2% BSA and 0.5% sodium azide in PBS). A Zeiss Axioskop 2 microscope was used to examine slides, and photos were taken using a Zeiss Axiocam camera and the associated Axiocam software. Images were processed in Adobe Photoshop; minimal adjustments were made, including rotation and resizing. Where image settings were altered, the adjustment was applied equally to affected and unaffected samples and to the whole image.

RNA extraction, RTPCR and qPCR

We collected the brains of adult *Espn*^{spdz} (P25) and *rthm* (P28) mice, which were snap-frozen in liquid nitrogen, and the organs of Corti of four-day-old (P4) *Atp2b2*^{Tkh} mice, which were dissected out and stored at -20°C in RNAlater stabilisation reagent (Ambion). All RNA dissections were carried out during a fixed time window to avoid circadian variation (*Espn*^{spdz}: between 2.5 and 3.5 hours after lights on; *rthm* and *Atp2b2*^{Tkh}: between 6 and 7.5 hours after lights on). For the brains, RNA was extracted using TRIzol, in some cases followed by processing through Direct-zol minipreps (Zymo Research, cat. no R2050). For the organs of Corti, RNA was extracted using either QIAshredder columns (QIAgen, cat. no. 79654) and the RNeasy mini kit (QIAgen, cat. no. 74104), or the Lexogen SPLIT kit (Lexogen, cat. no. 008.48), following the manufacturer's instructions. RNA concentration was measured using a Nanodrop spectrophotometer (ND-8000). RNA was normalised to the same concentration within each litter, then treated with DNAse 1 (Sigma, cat. no: AMPD1) before cDNA creation. cDNA was made using Superscript II Reverse Transcriptase (Invitrogen, cat. no: 11904-018) or Precision Reverse Transcription Premix (PrimerDesign, cat. no: RT-premix2). Primers for sequencing cDNA for testing *Espn* splicing in *Espn*^{spdz} mice were designed using Primer3[75] (Supplementary Table 3). Quantitative RT-PCR on cDNA from *rthm* and *Tkh* mice was carried out on a CFX Connect qPCR machine (Bio-Rad), using probes from Applied Biosystems (Hprt, cat. no: Mm01318747_g1; Jag1, cat. no: Mm01270190_m1; Atp2b2, cat. no: Mm01184578_m1; Slc12a2, cat no: Mm00436563_m1) and Sso-Advanced Master Mix (Bio-Rad, cat. no: 1725281). Relative expression levels were calculated using the $2^{-\Delta\Delta ct}$ equation[76], with *Hprt* as an internal control. *Jag1* was used to check for the quantity of sensory tissue present in the organ of Corti samples because it is expressed in supporting cells[77,78]. At least three technical replicates of each sample were carried out for each reaction. The Wilcoxon rank sum test (Mann-Whitney U test) was used to calculate significance for comparisons between *rthm* wildtypes and homozygotes. For the *Atp2b2*^{Tkh} qPCR, which involved three groups (wildtype, heterozygote and homozygote), a one-way ANOVA was carried out using SPSS (IBM).

References

1. Ayabe, S., Nakashima, K. & Yoshiki, A. Off- and on-target effects of genome editing in mouse embryos. *J Reprod Dev* **65**, 1-5 (2019).
2. Iyer, V. *et al.* No unexpected CRISPR-Cas9 off-target activity revealed by trio sequencing of gene-edited mice. *PLoS Genet* **14**, e1007503 (2018).
3. Willi, M., Smith, H.E., Wang, C., Liu, C. & Hennighausen, L. Mutation frequency is not increased in CRISPR-Cas9-edited mice. *Nat Methods* **15**, 756-758 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4. Anderson, K.R. *et al.* CRISPR off-target analysis in genetically engineered rats and mice. *Nat Methods* **15**, 512-514 (2018).
5. Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A. & Stambrook, P.J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci U S A* **99**, 3586-90 (2002).
6. Maitra, A. *et al.* Genomic alterations in cultured human embryonic stem cells. *Nat Genet* **37**, 1099-103 (2005).
7. Birling, M.C. *et al.* A resource of targeted mutant mouse lines for 5,061 genes. *Nat Genet* **53**, 416-419 (2021).
8. Skarnes, W.C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337-42 (2011).
9. White, J.K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452-64 (2013).
10. Bowl, M.R. *et al.* A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nat Commun* **8**, 886 (2017).
11. Ingham, N.J. *et al.* Mouse screen reveals multiple new genes underlying mouse and human hearing loss. *PLoS Biol* **17**, e3000194 (2019).
12. Ingham, N.J. *et al.* S1PR2 variants associated with auditory function in humans and endocochlear potential decline in mouse. *Sci Rep* **6**, 28964 (2016).
13. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-58 (2015).
14. Jiang, D.Q. *et al.* Crystal Structure Analysis of Kelch protein from Plasmodium falciparum. (PDB, 2015).
15. Murray, J.W. *et al.* Crystal structure of the BTB-BACK domains of human KLHL11. (PDB, 2009).
16. Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G. & Leprince, D. The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth Differ* **6**, 1193-8 (1995).
17. Bardwell, V.J. & Treisman, R. The POZ domain: a conserved protein-protein interaction motif. *Genes Dev* **8**, 1664-77 (1994).
18. Dhanoa, B.S., Cogliati, T., Satish, A.G., Bruford, E.A. & Friedman, J.S. Update on the Kelch-like (KLHL) gene family. *Hum Genomics* **7**, 13 (2013).
19. Ingham, N.J. *et al.* Inner hair cell dysfunction in Klhl18 mutant mice leads to low frequency progressive hearing loss. *bioRxiv*, 2021.03.09.434536 (2021).
20. Olesen, C. *et al.* The structural basis of calcium transport by the calcium pump. *Nature* **450**, 1036-42 (2007).
21. Carpinelli, M.R., Manning, M.G., Kile, B.T. & Burt, R.A. Two ENU-induced alleles of Atp2b2 cause deafness in mice. *PLoS One* **8**, e67479 (2013).
22. Takahashi, K. & Kitamura, K. A point mutation in a plasma membrane Ca(2+)-ATPase gene causes deafness in Wriggle Mouse Sagami. *Biochem Biophys Res Commun* **261**, 773-8 (1999).
23. Bortolozzi, M. *et al.* The novel PMCA2 pump mutation Tommy impairs cytosolic calcium clearance in hair cells and links to deafness in mice. *J Biol Chem* **285**, 37693-703 (2010).
24. Tsai, Y.S. *et al.* A de novo deafwaddler mutation of Pmca2 arising in ES cells and hitchhiking with a targeted modification of the Pparg gene. *Mamm Genome* **17**, 716-22 (2006).
25. Xu, L. *et al.* Identification of a novel point mutation of mouse Atp2b2 induced by N-ethyl-N-nitrosourea mutagenesis. *Exp Anim* **60**, 71-8 (2011).
26. Street, V.A., McKee-Johnson, J.W., Fonseca, R.C., Tempel, B.L. & Noben-Trauth, K. Mutations in a plasma membrane Ca2+-ATPase gene cause deafness in deafwaddler mice. *Nat Genet* **19**, 390-4 (1998).
27. Spiden, S.L. *et al.* The novel mouse mutation Oblivion inactivates the PMCA2 pump and causes progressive hearing loss. *PLoS Genet* **4**, e1000238 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

28. Komada, M., McLean, D.J., Griswold, M.D., Russell, L.D. & Soriano, P. E-MAP-115, encoding a microtubule-associated protein, is a retinoic acid-inducible gene required for spermatogenesis. *Genes Dev* **14**, 1332-42 (2000).
29. Magnan, D.R., Spacek, D.V., Ye, N., Lu, Y.C. & King, T.R. The male sterility and histoincompatibility (mshi) mutation in mice is a natural variant of microtubule-associated protein 7 (Mtap7). *Mol Genet Metab* **97**, 155-62 (2009).
30. Tian, C. & Johnson, K.R. TBX1 is required for normal stria vascularis and semicircular canal development. *Dev Biol* **457**, 91-103 (2020).
31. Lindsay, E.A. *et al.* Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* **410**, 97-101 (2001).
32. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-3 (2010).
33. Lefevre, G. *et al.* A core cochlear phenotype in USH1 mouse mutants implicates fibrous links of the hair bundle in its cohesion, orientation and differential growth. *Development* **135**, 1427-37 (2008).
34. Pawlowski, K.S., Kikkawa, Y.S., Wright, C.G. & Alagramam, K.N. Progression of inner ear pathology in Ames waltzer mice and the role of protocadherin 15 in hair cell development. *J Assoc Res Otolaryngol* **7**, 83-94 (2006).
35. Dixon, M.J. *et al.* Mutation of the Na-K-Cl co-transporter gene Slc12a2 results in deafness in mice. *Hum Mol Genet* **8**, 1579-84 (1999).
36. Bock, G.R. & Steel, K.P. Inner ear pathology in the deafness mutant mouse. *Acta Otolaryngol* **96**, 39-47 (1983).
37. Rzadzinska, A., Schneider, M., Noben-Trauth, K., Bartles, J.R. & Kachar, B. Balanced levels of Espin are critical for stereociliary growth and length maintenance. *Cell Motil Cytoskeleton* **62**, 157-65 (2005).
38. Sjostrom, B. & Anniko, M. Cochlear structure and function in a recessive type of genetically induced inner ear degeneration. *ORL J Otorhinolaryngol Relat Spec* **54**, 220-8 (1992).
39. Pettitt, S.J. *et al.* Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Methods* **6**, 493-5 (2009).
40. Herr, D.R. *et al.* Sphingosine 1-phosphate (S1P) signaling is required for maintenance of hair cells mainly via activation of S1P2. *J Neurosci* **27**, 1474-8 (2007).
41. Kono, M. *et al.* Deafness and stria vascularis defects in S1P2 receptor-null mice. *J Biol Chem* **282**, 10690-6 (2007).
42. MacLennan, A.J. *et al.* The S1P2 sphingosine 1-phosphate receptor is essential for auditory and vestibular function. *Hear Res* **220**, 38-48 (2006).
43. Watson, C.J. & Tempel, B.L. A new Atp2b2 deafwaddler allele, dfw(i5), interacts strongly with Cdh23 and other auditory modifiers. *Hear Res* **304**, 41-8 (2013).
44. Jerome, L.A. & Papaioannou, V.E. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nat Genet* **27**, 286-91 (2001).
45. Liao, J. *et al.* Full spectrum of malformations in velo-cardio-facial syndrome/DiGeorge syndrome mouse models by altering Tbx1 dosage. *Hum Mol Genet* **13**, 1577-85 (2004).
46. Raft, S., Nowotschin, S., Liao, J. & Morrow, B.E. Suppression of neural fate and control of inner ear morphogenesis by Tbx1. *Development* **131**, 1801-12 (2004).
47. Vitelli, F. *et al.* TBX1 is required for inner ear morphogenesis. *Hum Mol Genet* **12**, 2041-8 (2003).
48. Alagramam, K.N. *et al.* The mouse Ames waltzer hearing-loss mutant is caused by mutation of Pcdh15, a novel protocadherin gene. *Nat Genet* **27**, 99-102 (2001).
49. Zheng, L. *et al.* The deaf jerker mouse has a mutation in the gene encoding the espin actin-bundling proteins of hair cell stereocilia and lacks espins. *Cell* **102**, 377-85 (2000).
50. Zhang, Z. & Baldini, A. In vivo response to high-resolution variation of Tbx1 mRNA dosage. *Hum Mol Genet* **17**, 150-7 (2008).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

51. Chen, J. *et al.* Identification of a Novel ENU-Induced Mutation in Mouse Tbx1 Linked to Human DiGeorge Syndrome. *Neural Plast* **2016**, 5836143 (2016).
52. McCullough, B.J. & Tempel, B.L. Haplo-insufficiency revealed in deafwaddler mice when tested for hearing loss and ataxia. *Hear Res* **195**, 90-102 (2004).
53. Chen, Z., Hayasaka, S., Takagishi, Y., Murata, Y. & Oda, S. A novel mutant mouse, joggle, with inherited ataxia. *Exp Anim* **55**, 411-4 (2006).
54. Kozel, P.J. *et al.* Balance and hearing deficits in mice with a null mutation in the gene encoding plasma membrane Ca2+-ATPase isoform 2. *J Biol Chem* **273**, 18693-6 (1998).
55. Kumar, R.A. *et al.* Unexpected embryonic stem (ES) cell mutations represent a concern in gene targeting: lessons from "fierce" mice. *Genesis* **38**, 51-7 (2004).
56. Westrick, R.J. *et al.* Spontaneous Irs1 passenger mutation linked to a gene-targeted SerpinB2 allele. *Proc Natl Acad Sci U S A* **107**, 16904-9 (2010).
57. Ingham, N.J. Evoked Potential Recordings of Auditory Brainstem Activity in the Mouse: An Optimized Method for the Assessment of Hearing Function of Mice. *Bio-protocol* **9**, e3447 (2019).
58. Ingham, N.J., Pearson, S. & Steel, K.P. Using the Auditory Brainstem Response (ABR) to Determine Sensitivity of Hearing in Mutant Mice. *Curr Protoc Mouse Biol* **1**, 279-87 (2011).
59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
60. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
61. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
62. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-60 (2015).
63. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94 (2011).
64. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
65. The Broad Institute. Picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. Vol. 2016.
66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
67. Albers, C.A. *et al.* Dindel: accurate indel calls from short-read data. *Genome Res* **21**, 961-73 (2011).
68. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6**, 677-81 (2009).
69. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-71 (2009).
70. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
71. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
72. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
73. Bonfield, J.K., Smith, K. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res* **23**, 4992-9 (1995).
74. Hunter-Duvar, I.M. A technique for preparation of cochlear specimens for assessment with the scanning electron microscope. *Acta Otolaryngol Suppl* **351**, 3-23 (1978).
75. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).
76. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408 (2001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

77. Morrison, A., Hodgetts, C., Gossler, A., Hrabe de Angelis, M. & Lewis, J. Expression of Delta1 and Serrate1 (Jagged1) in the mouse inner ear. *Mech Dev* **84**, 169-72 (1999).
78. Zine, A., Van De Water, T.R. & de Ribaupierre, F. Notch signaling regulates the pattern of auditory hair cell differentiation in mammals. *Development* **127**, 3373-83 (2000).
79. El Omari, K. *et al.* Structure of the DNA-bound T-box domain of human TBX1, a transcription factor associated with the DiGeorge syndrome. *Proteins* **80**, 655-60 (2012).
80. Hoffmann, T.J. *et al.* A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. *PLOS Genetics* **12**, e1006371 (2016).
81. Wells, H.R.R. *et al.* GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *Am J Hum Genet* **105**, 788-802 (2019).
82. Hunt, S.E. *et al.* Ensembl variation resources. *Database (Oxford)* **2018**(2018).
83. Zerbino, D.R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

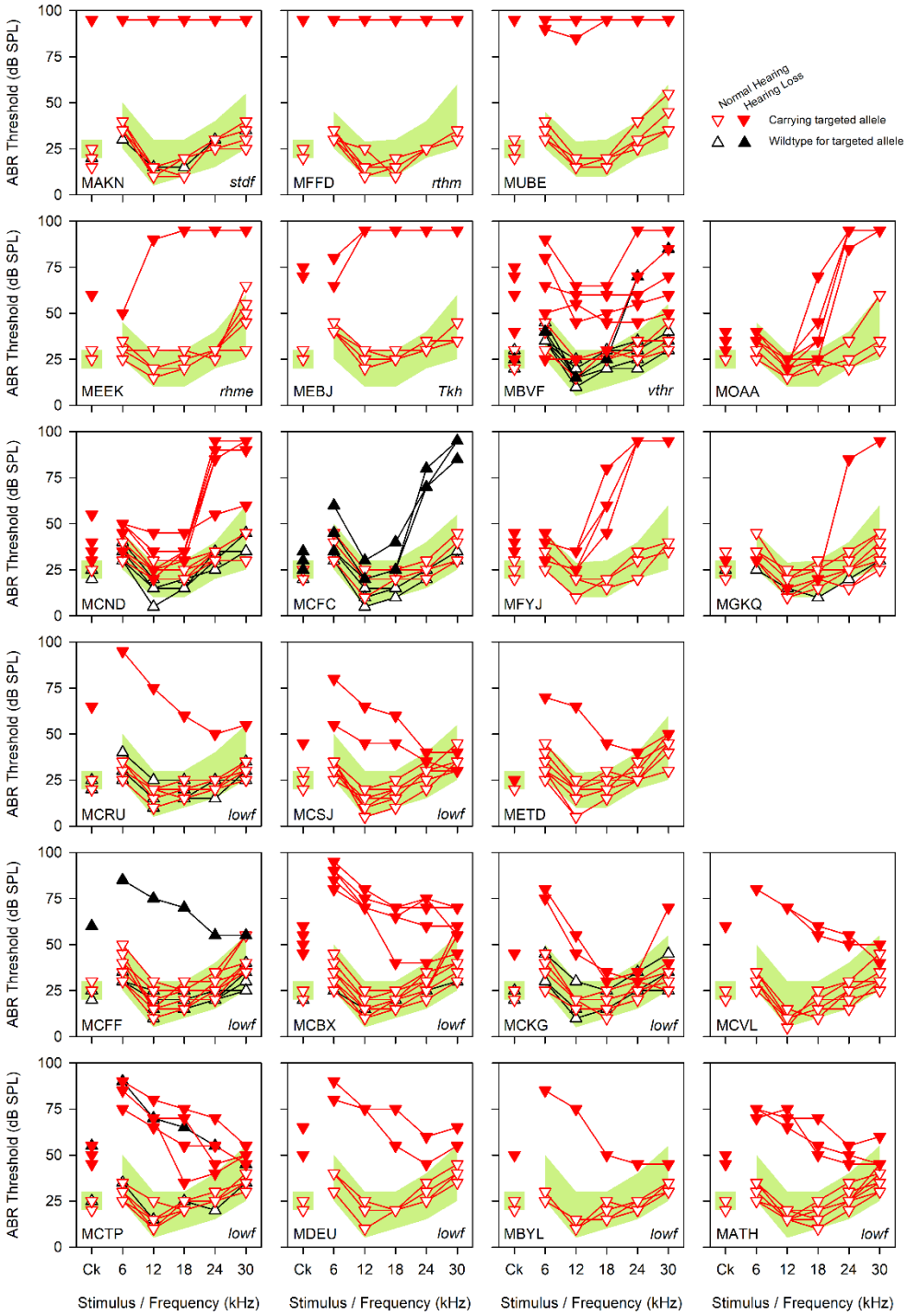


Fig. 1. ABR thresholds from 22 targeted lines displaying non-segregating hearing loss.

ABR thresholds showing individual mice from targeted mutant lines displaying hearing loss (filled symbols) which does not segregate with the targeted mutant allele (red, inverted triangles) from the original ABR screen. Mice with normal hearing are shown with empty symbols, and black triangles indicate mice which were wildtype for the targeted mutant allele. Affected mice from the MAKN, MFFD and MUBE lines (top) were almost completely deaf. Affected mice from the MEEK and MEBJ colonies exhibited severe hearing loss, while affected mice from the MBVF colony exhibited varying

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

degrees of hearing loss across all frequencies. Affected mice from the MOAA, MCND, MCFC, MFYJ and MGQK colonies (second and third lines) had hearing loss affecting only the higher frequencies, while the remaining eleven colonies all exhibited hearing loss affecting primarily the low frequencies. The shaded pale green area on each panel denotes the 95% reference range for a large population of control wildtype mice derived from littermates of the tested mice, defined in [11]. All mice were tested at 14 weeks old. Mice from three other lines which exhibited non-segregating hearing loss were detected because of their vestibular phenotype (due to the *Tbx1*^{ttch}, *Pcdh15*^{jigl} and *Espn*^{spdz} alleles), and did not undergo ABRs as part of the ABR screening pipeline. MAKN n=7 unaffected, 3 affected; MFFD n=5 unaffected, 1 affected; MUBE n=4 unaffected, 2 affected; MEEK n=5 unaffected, 1 affected; MEBJ n=4 unaffected, 2 affected; MBVF n=11 unaffected, 6 affected; MOAA n=3 unaffected, 6 affected; MCND n=9 unaffected, 5 affected; MCFC n=7 unaffected, 3 affected; MFYJ n=3 unaffected, 4 affected; MGKQ n=9 unaffected, 1 affected; MCRU n=11 unaffected, 1 affected; MCSJ n=7 unaffected, 2 affected; METD n=7 unaffected, 1 affected; MDEU n=5 unaffected, 2 affected; MCFF n=16 unaffected, 1 affected; MCBX n=9 unaffected, 5 affected; MCKG n=8 unaffected, 2 affected; MCVL n=7 unaffected, 2 affected; MCTP n=7 unaffected, 5 affected; MBYL n=5 unaffected, 1 affected; MATH n=8 unaffected, 3 affected.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

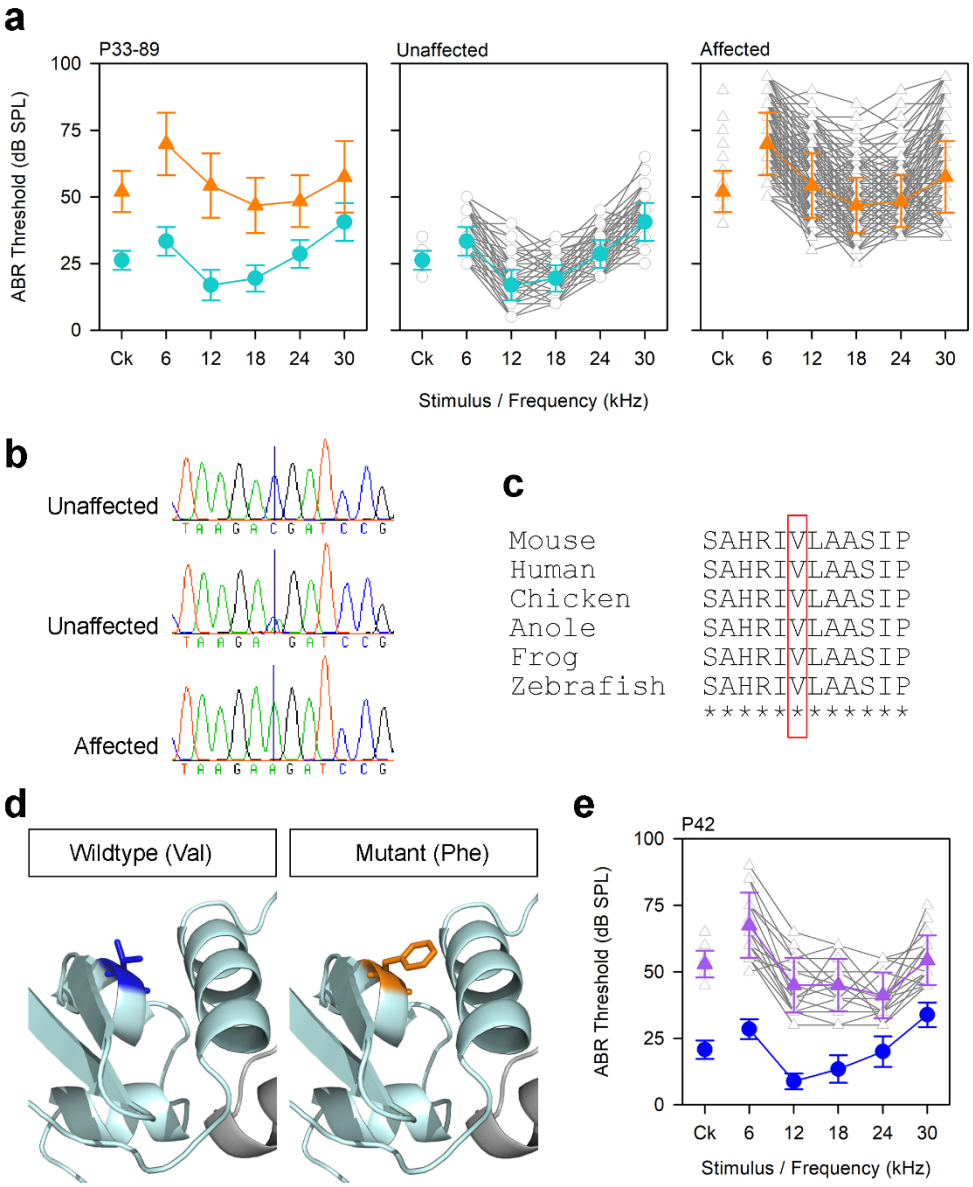


Fig. 2. Mutations in *Klhl18* cause low frequency hearing loss.

a ABR thresholds from mice tested during establishment of the breeding colony derived from the MCBX colony. Mice were tested between 33 and 89 days old, and grouped into affected (n=221, orange triangles) and unaffected (n=213, teal circles) based on the bimodal distributions of thresholds for clicks and 6kHz stimuli. Plots of individual ABR thresholds (grey) are shown separately with the mean trace indicated by coloured lines and symbols; error bars on mean trace are standard deviations. **b** Sequence traces from two unaffected mice (one wildtype, one heterozygote) and an affected mouse (homozygote) showing the variant *Klhl18*^{lowf} (MCBX colony), p.V55F. **c** Clustal alignment from mouse, human, chicken, anole lizard, frog and zebrafish showing that the affected amino acid is highly conserved (red box). **d** Close-up of the BTB domain (pale cyan) showing the amino acid structures for the wildtype residue (blue, left) and the mutant residue (orange, right). **e** ABR thresholds from mice heterozygous for the *lowf* allele (n=13, blue circles) and compound heterozygotes carrying the *lowf* allele and the *Klhl18*^{tm1a} allele (n=17, purple triangles), demonstrating the low frequency hearing loss phenotype. Individual traces for the compound heterozygotes are shown in grey; error bars on mean trace are standard deviations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

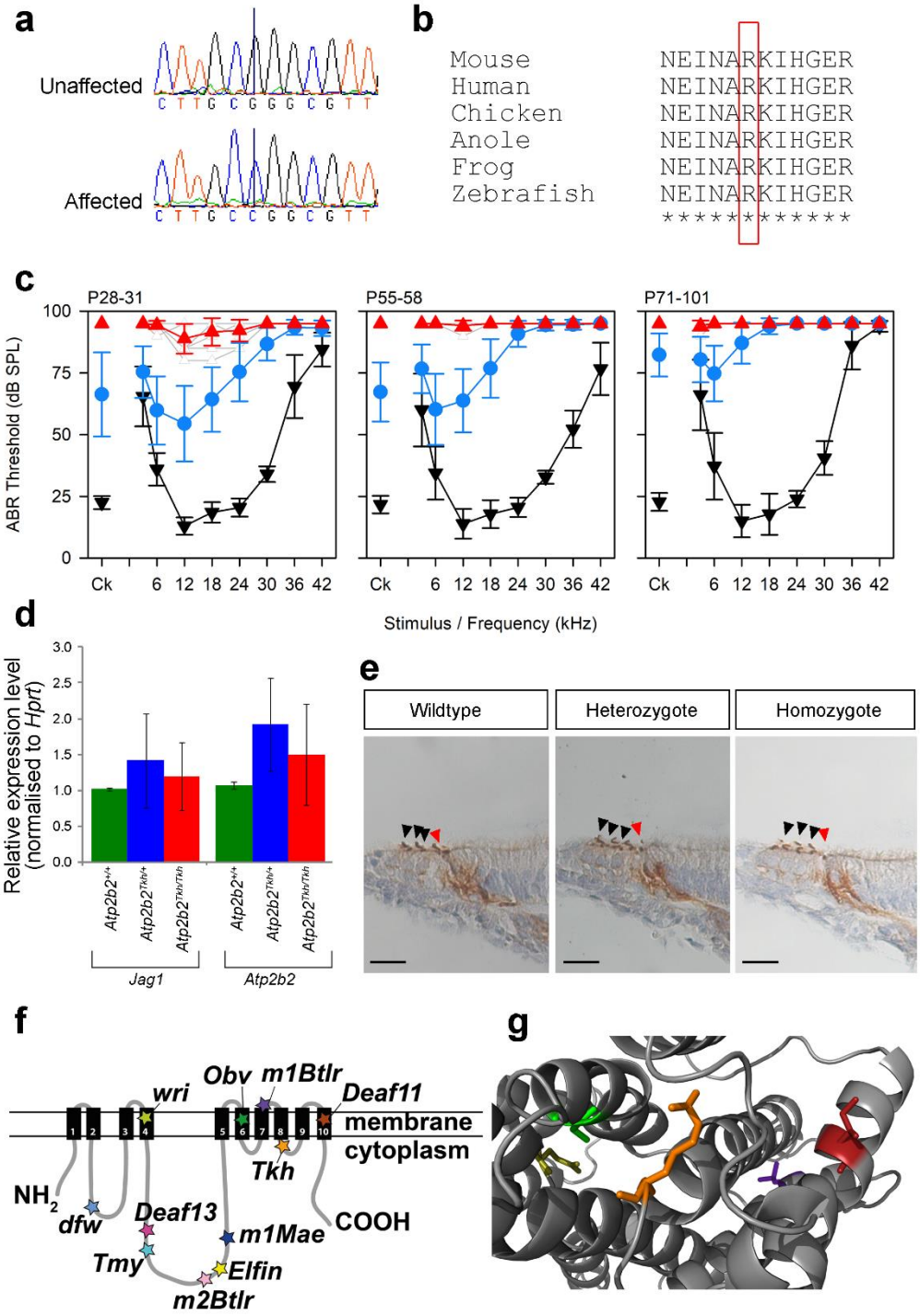


Fig. 3. A missense mutation in *Atp2b2* results in semidominant progressive hearing loss.

a Sequence traces from an unaffected and an affected mouse showing the variant *Atp2b2*^{Tkh} (MEBJ colony), p.R969G. **b** Clustal alignment from mouse, human, chicken, anole lizard, frog and zebrafish showing that the affected amino acid is highly conserved (red box). **c** ABR thresholds from wildtype (black inverted triangles), heterozygote (blue circles) and homozygote (red triangles) mice at P28-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

P31 (n=10 wildtypes, 27 heterozygotes, 9 homozygotes), P55-P58 (n= 9 wildtypes, 24 heterozygotes, 6 homozygotes) and P71-P101 (n=9 wildtypes, 24 heterozygotes, 6 homozygotes). **d** *Jag1* and *Atp2b2* qPCR on RNA from the organ of Corti at P4 (n = 6 wildtypes, 6 heterozygotes and 6 homozygotes). No significant difference was seen between wildtypes, heterozygotes and homozygotes (*Jag1* p=0.357, *Atp2b2* p=0.057, one way ANOVA). **e** PMCA2 antibody stains at P4 (n=2 wildtypes, 2 heterozygotes, 2 homozygotes), showing hair cells from the region 43% of the distance along the organ of Corti from base to apex. Images are representative examples for each genotype, and no differences were observed between genotypes. Arrowheads indicate the hair cells, red for the inner hair cell and black for the outer hair cells. Scale bar = 20µm. **f** Schematic of PMCA2 protein showing the 10 transmembrane helices and the 11 known missense mutations[21-27] ((https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=1638) and (https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=3670)). **g** Model of PMCA2 protein with the amino acid affected by the *Tkh* allele in orange. The four other missense mutations located in or near the transmembrane helices are also visible, shown in green (*Obv* [27]), dark red (*Deaf11* [21]), olive yellow (*wri* [22]) and purple (*M1Btlr*, https://mutagenetix.utsouthwestern.edu/phenotypic/phenotypic_rec.cfm?pk=1638). The wildtype amino acid structures are shown for all four residues.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

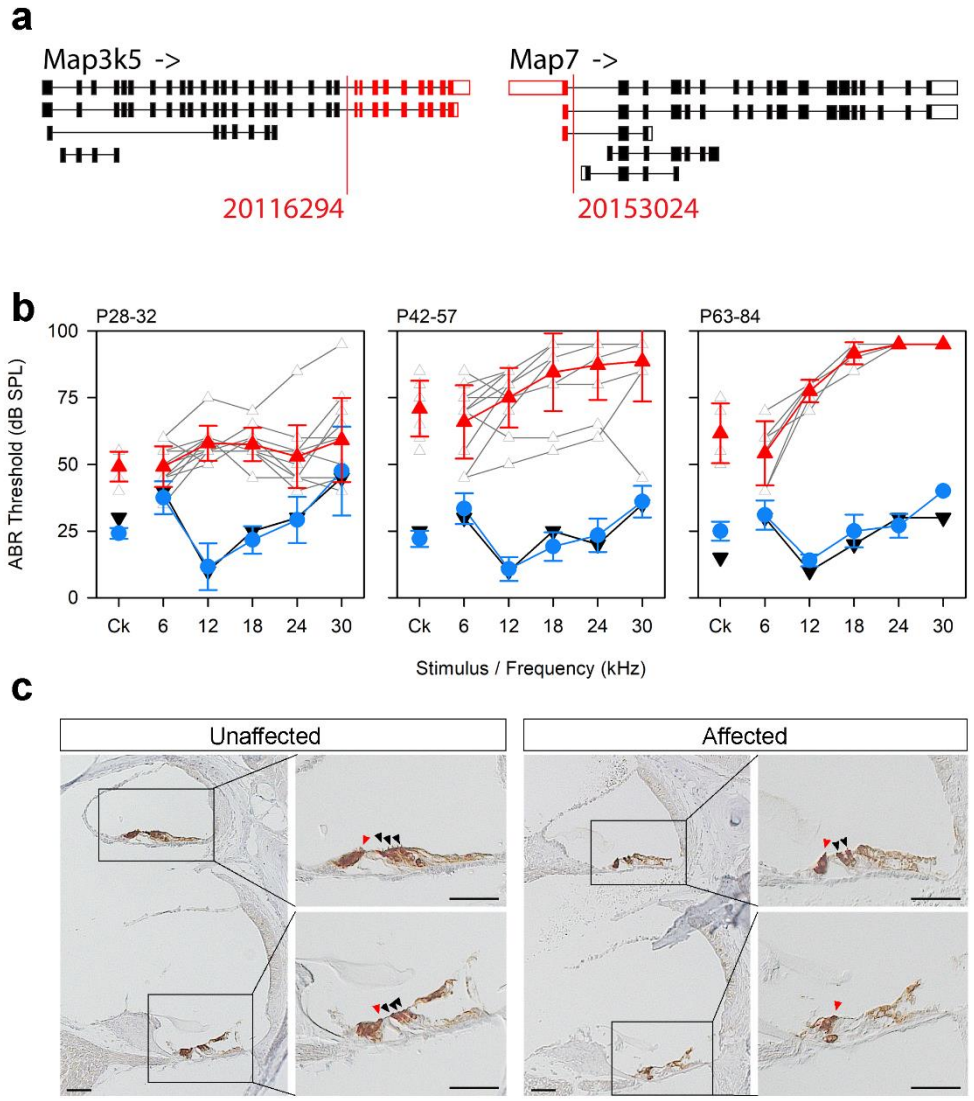


Fig. 4. The *rhme* deletion spanning *Map3k5* and *Map7* causes progressive hearing loss and disrupts the organ of Corti.

a Schematic of the *rhme* deletion seen in the MEEK line showing the different isoforms of the two affected genes (not to scale). Red indicates missing exons. **b** ABR thresholds showing progressive hearing loss in mice homozygous for the deletion (wildtype shown by black inverted triangles, heterozygote by blue circles, homozygotes by red triangles) (n = 1 wildtype, 6 heterozygotes and 12 homozygotes at P28-32; 1 wildtype, 19 heterozygotes and 11 homozygotes at P42-57; 1 wildtype, 5 heterozygotes and 6 homozygotes at P63-84). Traces from individual homozygotes are shown in grey. **c** Expression of MYO7A, a hair cell marker, in the organ of Corti of adult mice showing an apical turn (top; 94% of the distance along the organ of Corti from base to apex), where the hair cells are still present in affected mice, and a mid-basal turn (bottom, 43% of the distance from base to apex), where only the inner hair cell is identifiable (hair cells are marked with arrowheads; red for inner hair cells, black for outer hair cells). Brown indicates where MYO7A is expressed. Three affected and 3 unaffected littermates (at matched ages between 33 and 84 days old) were examined, and the images shown are representative of our observations. Scale bar = 50µm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

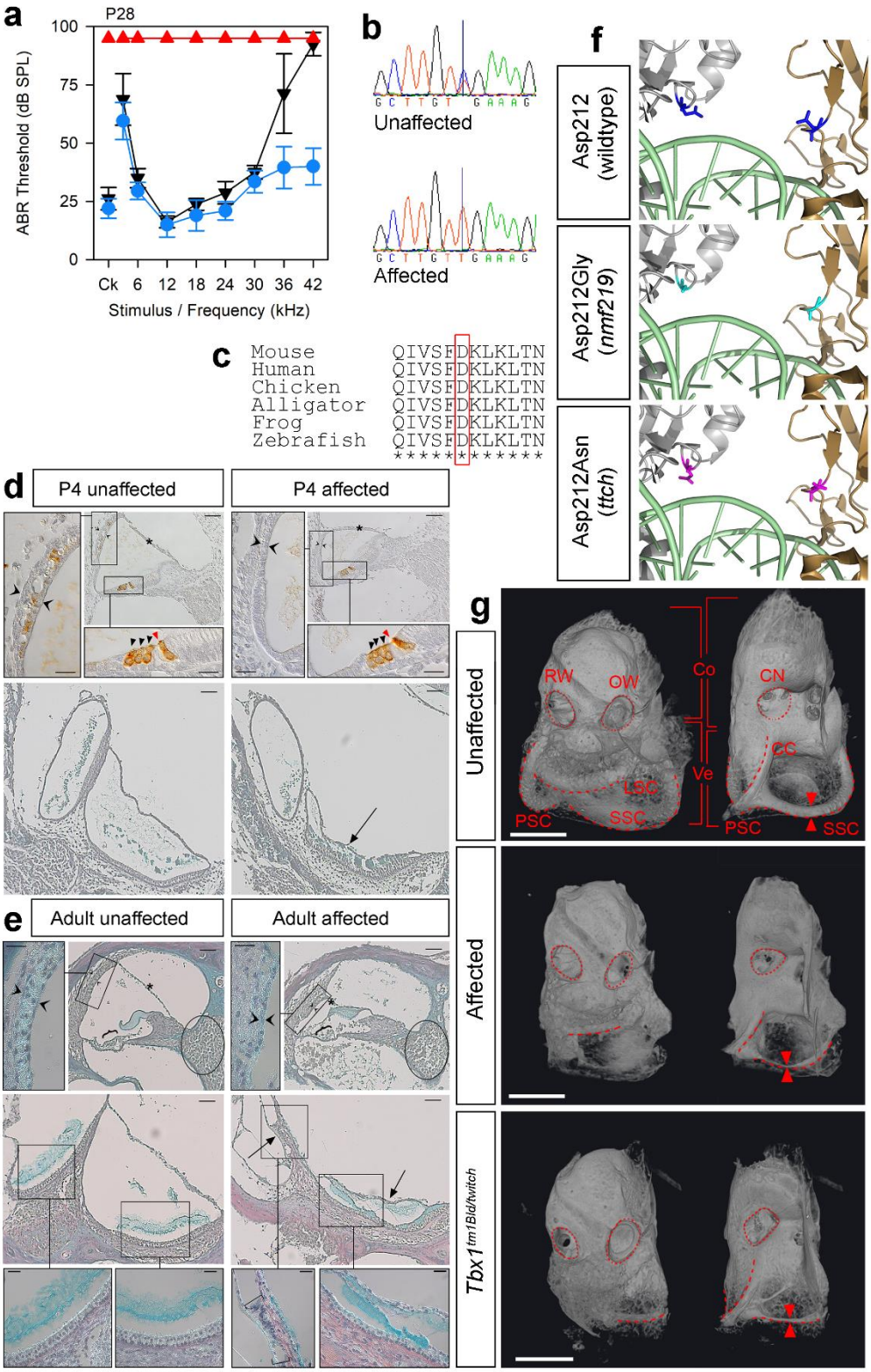


Fig. 5. A missense mutation in *Tbx1* causes malformation of the semicircular canals and profound deafness.

a ABR thresholds from P28 mice homozygous (n=6, red triangles), heterozygous (n=10, blue circles) and wildtype (n=4, black inverted triangles) for the *Tbx1*^{ttch} allele (MDLY colony), p.D212N. **b** Sequence traces from an unaffected and an affected mouse showing the variant. **c** Clustal alignment showing conservation of the affected amino acid (red box). **d** P4 sections showing the cochlear duct

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(top, anti-MYO7A brown stain, blue counterstain, 43% of the distance along the organ of Corti from base to apex) and the maculae (bottom, trichrome staining) (n=3 affected, 3 unaffected littermates). MYO7A is expressed in hair cells (arrowheads; red/black for inner/outer hair cells) and the intermediate cells of the stria vascularis. **e** Trichrome-stained sections from adult mice (n=4 affected, 4 unaffected littermates) showing the cochlear duct (top, 72% of the distance along the organ of Corti from base to apex) and the maculae (bottom). Brackets mark the organ of Corti and the spiral ganglion area is circled. Square brackets indicate abnormal saccular hair cells. In **d** and **e**, main panel scale bars = 50µm, high magnification panel scale bars = 20µm. Asterisks mark Reissner's membrane, twin open arrowheads the stria vascularis, and arrows the collapsed saccule. **f** Human TBX1 protein model[79], as a homodimer (silver, gold) bound to DNA (pale green). The Asp212 residue is marked in blue (top), with the *nmf219* mutant residue[30] in cyan (middle) and the *ttch* mutant residue in magenta (bottom). **g** MicroCT scans of cleared inner ears from affected P28 mice (n=3, middle) and compound heterozygotes ($Tbx1^{tm1Bld/ttch}$) at P21 (n=2, bottom). An unaffected P21 mouse is shown at the top (P21 n=4; P28 n=3). Dashed lines outline the semicircular canals, with twin arrowheads for comparison of their width. The middle ear side, with the round (RW) and oval (OW) windows (dotted lines) is on the left, and the brain side, where the cochlear nerve exits (CN, dotted lines), on the right. Brackets indicate the cochlea (Co) and vestibular region (Ve). LSC=lateral semicircular canal; SSC=superior semicircular canal; PSC=posterior semicircular canal, CC=common crus. Scale bar = 1mm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

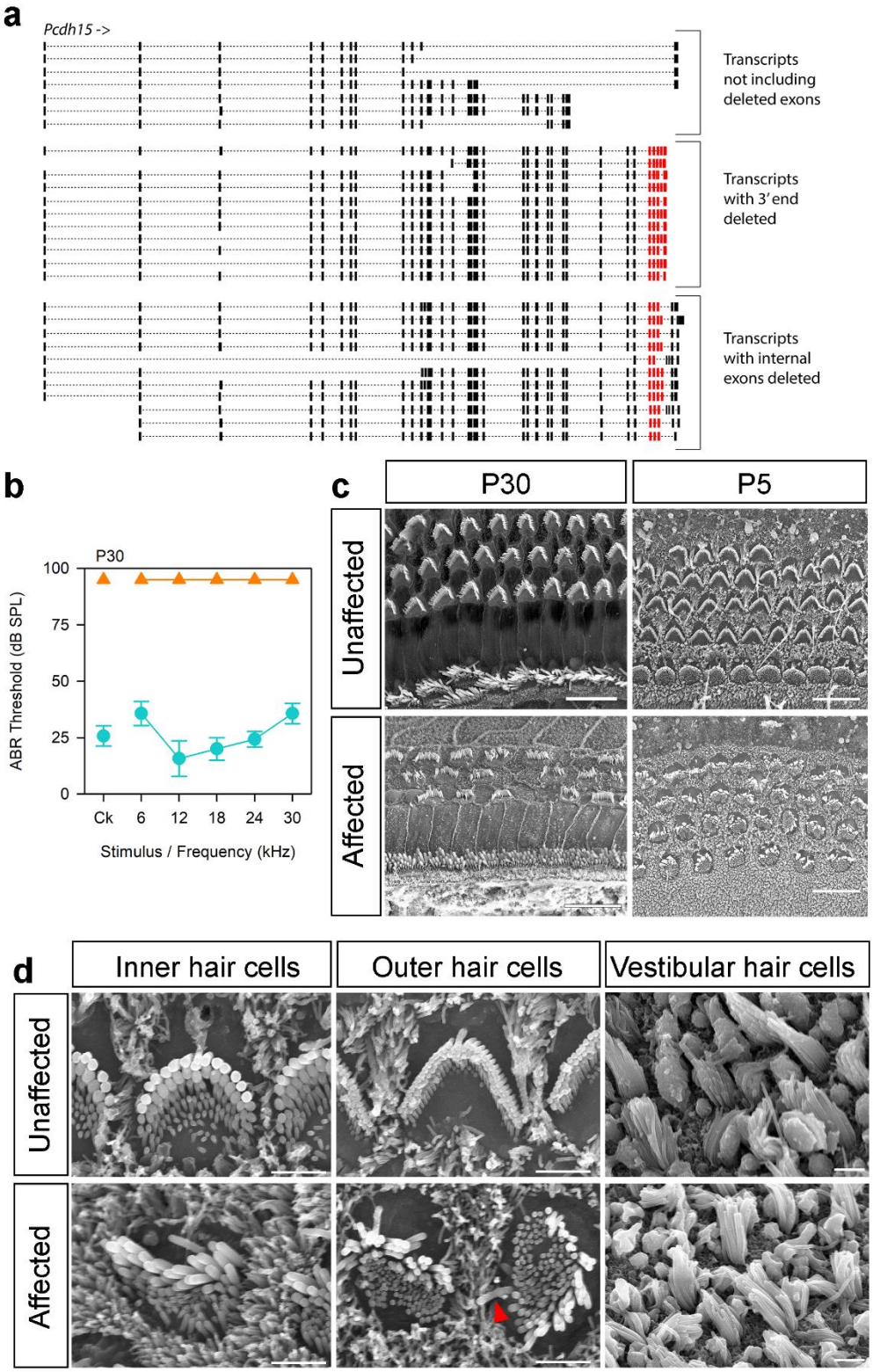


Fig. 6. A deletion within the *Pcdh15* gene causes disruption of stereocilia bundles and profound deafness.

a Schematic of *Pcdh15* isoforms showing the internal deletion in *Pcdh15* (*jigl*, MEWY colony*)*, showing the effects of the deletion on different transcripts (not to scale). Red indicates the deleted exons. Transcripts are shown in their entirety, but not all exons are visible due to scale. **b** ABR thresholds of affected (n=6, orange triangles) and unaffected (n=7, teal circles) mice at P30. **c**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Scanning electron micrographs of the organ of Corti at P30 (60-70% of the distance from base to apex, scale bar = 10µm, n=3 unaffected, 6 affected mice) and P5 (50-80% of the distance from base to apex, scale bar = 10µm, n=5 unaffected, 4 affected mice). Mice at P5 were not able to be phenotyped by circling behaviour, but the scanning electron micrographs showed a clear bimodal distribution of affected and unaffected based on disorganisation of hair bundles, and we were unable to amplify the deleted region in affected pups (see Table S7 for primers). Representative examples are shown here. Affected P5 mice had less well-organised hair cell bundles than unaffected littermates. The overall "V" shape of the stereocilia bundle was distorted and irregular in the organ of Corti of affected mice. The polarity of some outer hair cell bundles was rotated by up to 90°, and in some cases the kinocilium was on the opposite side of the cell (arrowhead in **d**). **d** Close ups of hair cells at P5, showing cochlear hair cells from the same region as in (c), and vestibular hair cells from the macula (n= 2 unaffected and 2 affected mice). The arrowhead indicates an example of a kinocilium on the opposite side of the outer hair cell from the stereocilia bundle. Macular stereocilia bundles seemed to have a more ordered staircase structure in the mice classed as unaffected compared to affected mice, where all stereocilia appeared to be long with little sign of a staircase arrangement. Scale bars = 2µm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

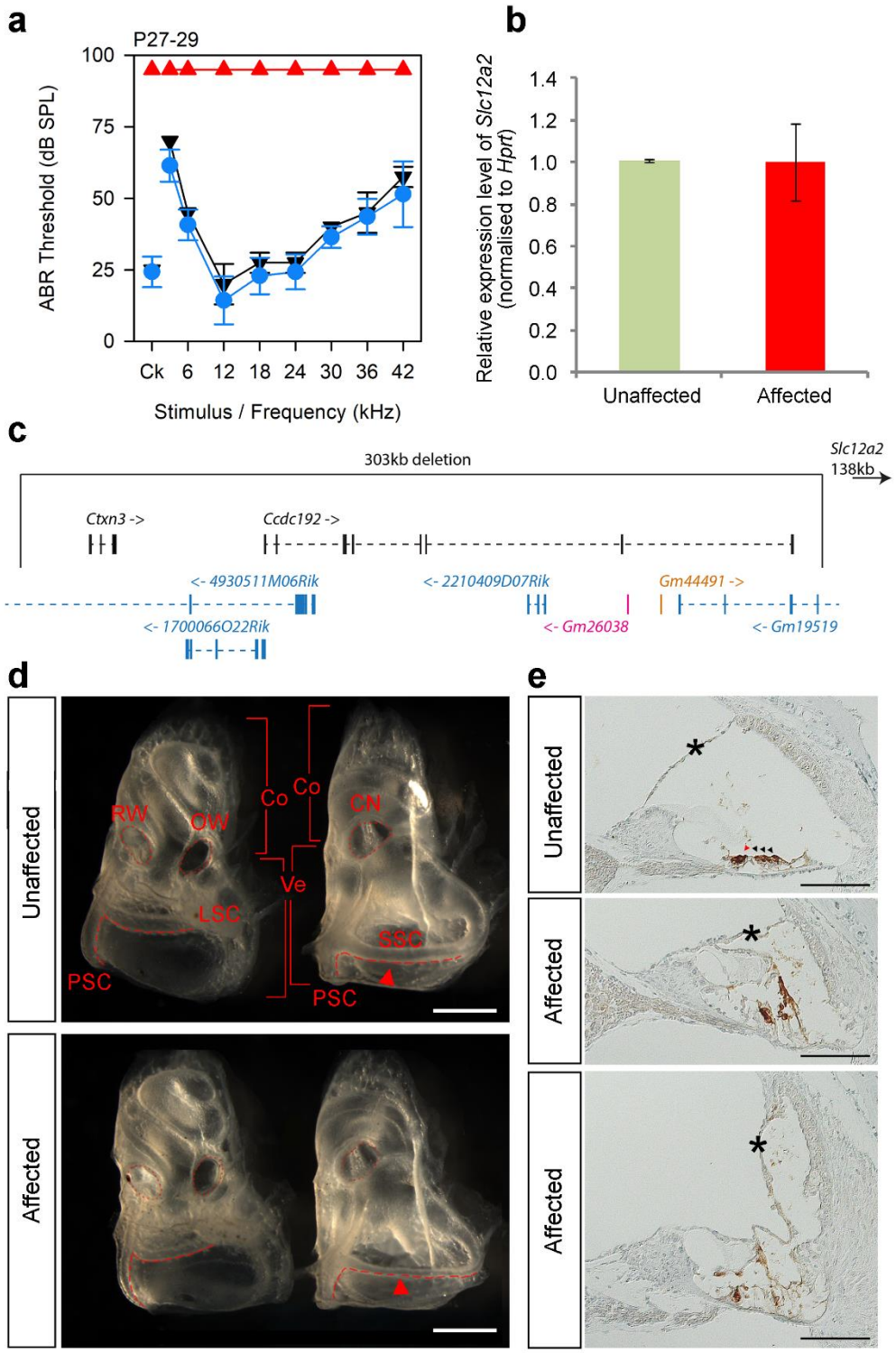


Fig. 7. A 303kb deletion on chr18 affects semicircular canals and results in profound deafness.

a ABR thresholds of mice homozygous (n=6, red triangles), heterozygous (n=7, blue circles) and wildtype (n=2, black inverted triangles) for the *rthm* allele (MFFD colony) at P28±1 day. **b** *Slc12a2* qPCR on RNA from the brains of P28 affected (n=4, red, right) and unaffected (n=4, green, left) mice. There is no significant difference between wildtypes and homozygotes (p=1, Wilcoxon rank sum test). **c** Schematic showing the genes affected by the deletion. There are two protein-coding genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(black), one miRNA gene (orange), one snRNA gene (pink) and four lncRNA genes (blue). *Slc12a2* is located 138kb downstream of the 3' end of the deletion (indicated by arrow top right). **d** Inner ears from affected (n=3) and unaffected (n=3) mice at P28±1 day. The middle ear side, with the round (RW) and oval (OW) windows is shown on the left of each panel, and the brain side, where the cochlear nerve exits (CN), on the right. The round and oval windows and the cochlear nerve exits are marked by dotted lines, and the semicircular canals by dashed lines. Brackets indicate the cochlea (Co) and vestibular region (Ve). LSC=lateral semicircular canal; SSC=superior semicircular canal; PSC=posterior semicircular canal. Scale bar = 1mm.The superior semicircular canals are thinner in affected mice (red arrowheads). **e** Immunohistochemistry of the cochlear duct of affected (n=3) and unaffected (n=3) mice at P28. Brown indicates the presence of MYO7A, which marks hair cells and the intermediate cells of the stria vascularis. Hair cells are only clearly visible in the unaffected mouse, indicated by the arrowheads (red for the inner hair cell and black for the outer hair cells). Asterisks show the Reissner's membrane, which is displaced in affected mice. The top two panels show the apical region, 72% of the distance from the base to the apex. The lower panel shows the basal region, at 16% of the base-apex distance. Scale bar = 100mm.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

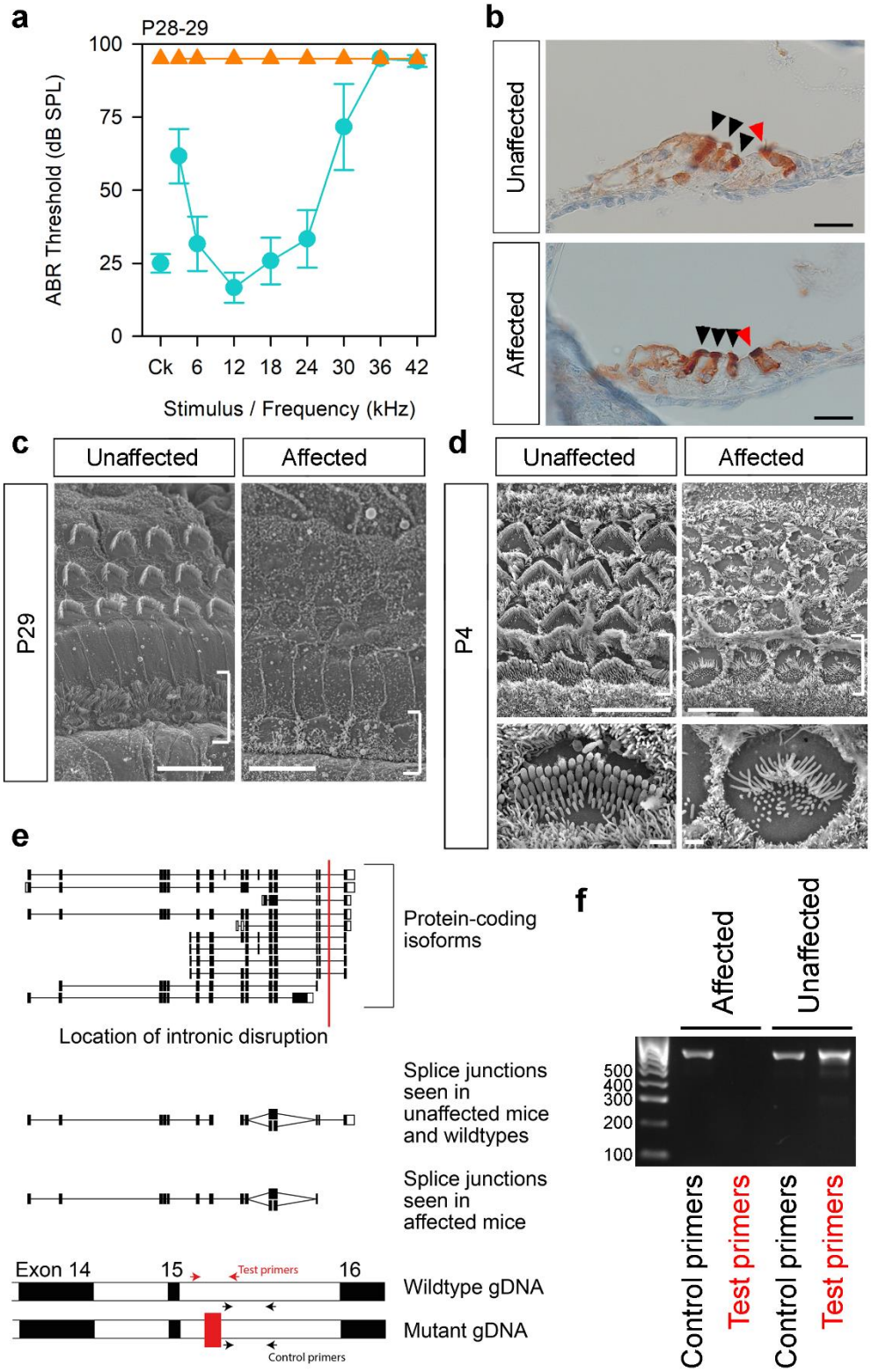


Fig. 8. Mice carrying the *spdz* allele (MHER colony) are profoundly deaf, with abnormal hair cell stereocilia at P5 and hair cell degeneration by P29.

a ABR thresholds of affected (orange triangles, n=7) and unaffected (teal circles, n=6) mice at P28-29. **b** Expression of MYO7A (brown) in hair cells from the apical region of the organ of Corti (90-95% of the distance from base to apex) of mice at P28 (n=3 affected, 3 unaffected), showing that hair cells are visible in affected mice (arrowheads, red for the inner hair cell and black for outer hair

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cells). Scale bar = 20µm. **c** Scanning electron micrographs of the 12kHz best-frequency region (65-70% of the distance from base to apex) of the organ of Corti in unaffected (n=4) and affected (n=3) mice at P29. Brackets indicate the inner hair cell row. Most hair bundles are missing. Scale bar = 10µm. **d** Scanning electron micrographs of a homozygous mutant mouse at P4 (n=3) and an age-matched wildtype, showing the 42kHz best-frequency region (20% of the distance from base to apex). Brackets indicate the inner hair cell row; scale bar = 10µm. Lower panels show a magnified view of inner hair cells in the same region (scale bar = 1µm). **e** Schematic showing the known protein-coding isoforms of *Espn* and the location of the intronic disruption (red line). The splice junctions identified in brain cDNA are shown below; each line indicates a splice junction visible in at least one mouse. At the bottom is a schematic of the genomic DNA (gDNA) showing exons 14, 15 and the start of exon 16 in black, separated by introns in white (not to scale). The disruption in the mutant is shown by the addition of a red block in the intron. The “test” primer pair (red arrows) fails to amplify in mutant gDNA, and the “control” primer pair (black arrows) works for both wildtype and mutant. **f** Gel showing bands amplified using these primer pairs (Supplementary Table 3) on an affected mouse and an unaffected mouse. The size of ladder bands is shown at the side in bp.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

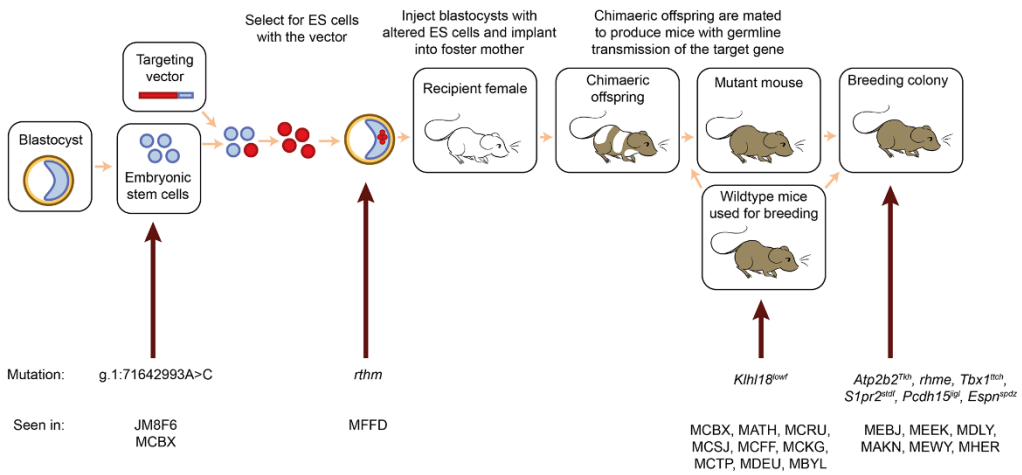


Fig. 9. There are many opportunities for spontaneous mutations to arise during the process of making a targeted knockout allele.

A schematic showing the stages of making knockout mice. Spontaneous mutations can arise at any point in this process. The mutations described in this paper are shown at the bottom, with arrows indicating the latest possible time at which that mutation could have occurred. If a spontaneous mutation occurs in the cultured embryonic stem cells before targeting, it has the potential to affect multiple mouse lines, but although we detected variants in MCBX mice which were present in the parental ES cell line JM8F6 (eg g.1: 71642993A>C, Supplementary Fig 6, Supplementary Table 4b), none of the eight mutations affecting hearing were found in any of the parental ES cell lines (Table 1). A mutation arising later in the process may be specific to a single line (such as the *rthm* allele, which is likely to have arisen in the targeted ES cell) or a single mating within that line, which is likely to be the case for most of the mutations described here. The *Klhl18*^{lowf} mutation, on the other hand, probably arose within the wildtype line used for expansion for all the colonies in which mice with the low frequency hearing loss phenotype were found.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Allele name	Allele symbol	Targeted gene	Line	Phenotype	Mutation	Affected gene(s)	MGI Allele symbol	MGI ID	Mutation type	Mutation effect	Inheritance	Parental ES Cell line	Mutation present in parental ES cell line	Mutation type compared to other reported mutant alleles in gene
Stone deaf	*stdf*	*Mms22l*	MAKN	Rapidly progressive severe hearing loss	g.9:20967665G>C	*S1pr2*	*S1pr2<stdf>*	MGI:5423977	Missense	T289R	Recessive	JM8.N4	No	Similar
Tikho	*Tkh*	*Rasal2*	MEBJ	Progressive hearing loss affecting higher frequencies first	g.6:113759212G>C	*Atp2b2*	*Atp2b2<Tkh>*	MGI:6470863	Missense	R969G	Semi- dominant	JM8A3.N1	No	Hypomorph
Twitch	*ttch*	*Arpc3*	MDLY	Complete deafness with vestibular dysfunction	g.16:18584128C>T	*Tbx1*	*Tbx1<ttch>*	MGI:6470864	Missense	D212N	Recessive	JM8A3.N1	No	Hypomorph
Jiggle	*jigl*	*Ccdc122*	MEWY	Complete deafness with vestibular dysfunction	Deletion between g.10:74614441 and g.10:74635149	*Pcdh15*	*Pcdh15<jigl>*	MGI:6470865	Deletion, with potential insertion	Loss of four exons towards 3' end of transcript	Recessive	JM8A1.N3	No	Similar
Rhythm	*rthm*	*Isg20*	MFFD	Complete deafness with vestibular dysfunction	g.18:57437258_57740507del	*Ctxn3, Ccdc192 and 6 noncoding genes*	*Del(18Ctxn3-Ccdc192)1Kcl*	MGI:6470868	Deletion	303kb deletion	Recessive	JM8A3.N1	No	N/A
Rhyme	*rhme*	*Rhox13*	MEEK	Progressive hearing loss affecting higher frequencies first	g.10:20116294_20153024del	*Map3k5, Map7*	*Del(10Map3k5-Map7)2Kcl*	MGI:6470869	Deletion	36.7kb deletion affecting last 9 exons of Map3k5 and first exon of Map7	Recessive	JM8A3.N1	No	Male sterility of *Map7* mutants is the same; hearing loss has not been reported
Spindizzy	*spdz*	*Slc35f2*	MHER	Complete deafness with vestibular dysfunction	Disruption between g.4:152122586 and g.4:152123017	*Espn*	*Espn<spdz>*	MGI:6470866	Deletion, with potential insertion or other disruption	Disruption in intron	Recessive	JM8A3.N1	No	Similar
Low frequency	*lowf*	*Mab21l4*	MCBX	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.F6	No	Similar
Low frequency	*lowf*	*Traf3ip3*	MATH	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.N4	No	Similar
Low frequency	*lowf*	*Mad2l2*	MBYL	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.F6	No	Similar
Low frequency	*lowf*	*Kazn*	MCFF	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.N4	No	Similar
Low frequency	*lowf*	*Sesn3*	MCKG	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.N4	No	Similar
Low frequency	*lowf*	*Pabpc1l*	MCRU	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8A3.N1	No	Similar
Low frequency	*lowf*	*Kcne2*	MCSJ	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8.N4	No	Similar
Low frequency	*lowf*	*Ptpn2*	MCTP	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	JM8A3.N1	No	Similar
Low frequency	*lowf*	*Tnfaip1*	MDEU	Low frequency hearing loss	g.9:110455454C>A	*Klhl18*	*Klhl18<lowf>*	MGI:6470867	Missense	V55F	Recessive	Unknown (imported line)	Unknown	-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC BY 4.0 International license.

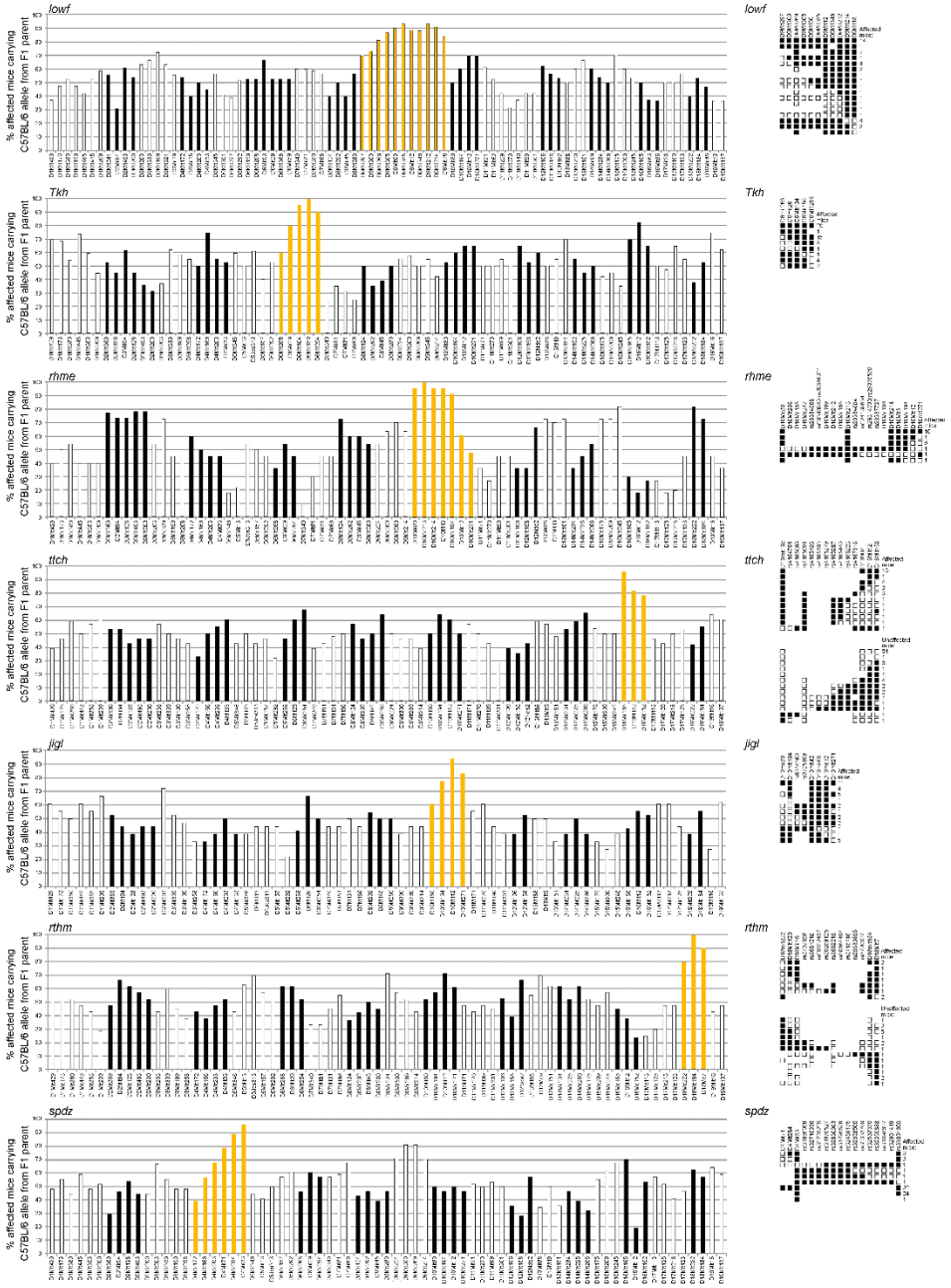
Variable thresholds	*vthr*	*Dusp3*	MBVF	Variable moderate to severe hearing loss across all frequencies	Unknown	Unknown	-	-	-	-	-	JM8.N19	Unknown	-
-	Un-known	*Rbcd3*	MBTD	Low frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8A1.N3	Unknown	-
-	Un-known	*Trim66*	MCVL	Low frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8.N4	Unknown	-
-	Un-known	*Tpi1*	MCFC	High-frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8.N4	Unknown	-
-	Un-known	*Aff3*	MCND	High-frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8.N4	Unknown	-
-	Un-known	*Wdtc1*	MFYJ	High-frequency hearing loss, only seen in females	Unknown	Unknown	-	-	-	-	-	JM8.N4	Unknown	-
-	Un-known	*Ddah1*	MGKQ	High-frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8A3.N1.C2	Unknown	-
-	Un-known	*Mir32*	MOAA	High-frequency hearing loss	Unknown	Unknown	-	-	-	-	-	JM8A3	Unknown	-
-	Un-known	*Sfxn3*	MUBE	Severe & profound hearing loss	Unknown	Unknown	-	-	-	-	-	JM8A3.N1	Unknown	-

Table 1. Summary of the 25 lines found to have an spontaneous mutation affecting hearing.

This table summarises the 25 lines in which hearing phenotypes were observed which did not segregate with the targeted allele. While some of the targeted genes have been associated with a hearing phenotype (eg *Sesn3* homozygotes have abnormal waveforms[11], and *ISG20* has been associated with human hearing loss via GWAS[80,81]), the targeted alleles did not segregate with the phenotypes reported in this paper, and are not linked to the observed hearing loss described here.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Figures and Tables



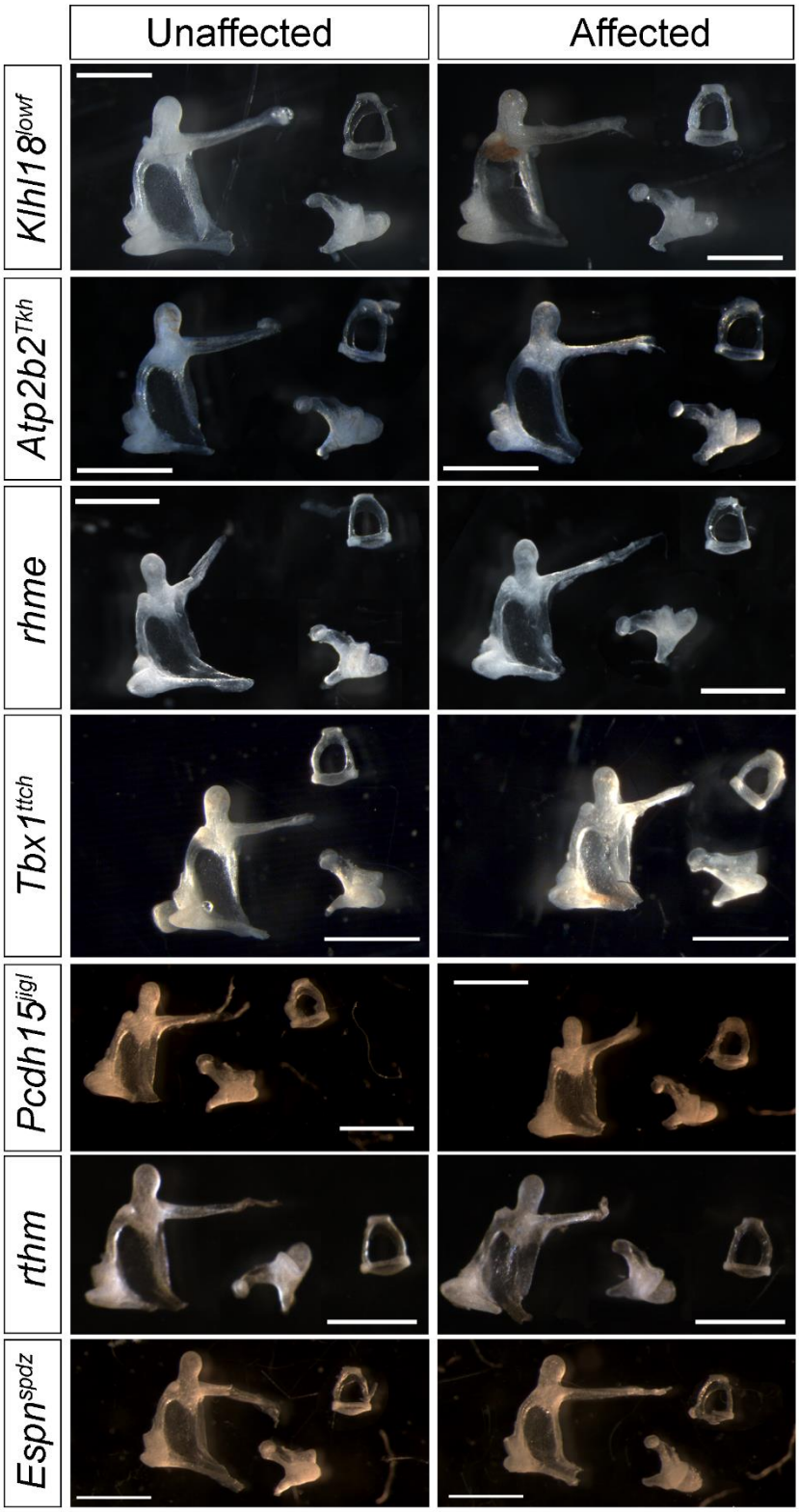
Supplementary Fig. 1. Genome scans and fine mapping.

The bar charts show the percentage of affected mice carrying the C57BL/6N allele on the chromosome derived from the outcrossed parent at each marker. Alternate chromosomes are indicated by different shading, and the linked chromosome for each mutation is coloured gold. Following the initial mapping, the critical region was identified using more markers and, where necessary, SNVs known to differ between C57BL/6N and C3HeB/FeJ, the two strains used for the mapping crosses. On the right, the haplotype charts show the fine mapping of each allele. Each row represents the marker pattern for the chromosome derived from the F1 parent, with white boxes representing the C3HeB/FeJ-like marker types and black boxes representing the C57BL/6N-like marker types. A change from black boxes to white or vice versa indicates that a recombination event has occurred during meiosis leading to the production of that chromosome in the F1 parent. For all

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

alleles except *Tkh*, the F1 parent was backcrossed to the affected parental line to obtain unaffected mice heterozygous for the causative mutation and affected homozygotes, but because the *Tkh* allele was semidominant, the F1 parent was backcrossed to C3HeB/FeJ, resulting in affected mice heterozygous for the causative mutation and unaffected wildtypes. The number to the right of each row indicates the number of mice showing that particular pattern of marker types. Numbers: *lowf*: n=30 affected mice for the genome scan, 44 affected mice for fine mapping; *Tkh*: n = 20 affected mice for the genome scan, 61 affected mice for fine mapping; *rhme*: n = 11 affected mice for the genome scan, 23 affected mice for fine mapping; *ttch*: n = 28 affected mice for the genome scan, 31 affected and 75 unaffected mice for fine mapping; *jigl*: n = 18 affected mice for the genome scan, 35 affected mice for fine mapping; *rthm*: n = 21 affected mice for the genome scan, 9 affected and 19 unaffected mice for fine mapping; *spdz*: n = 28 affected mice for the genome scan, 56 affected mice for fine mapping.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



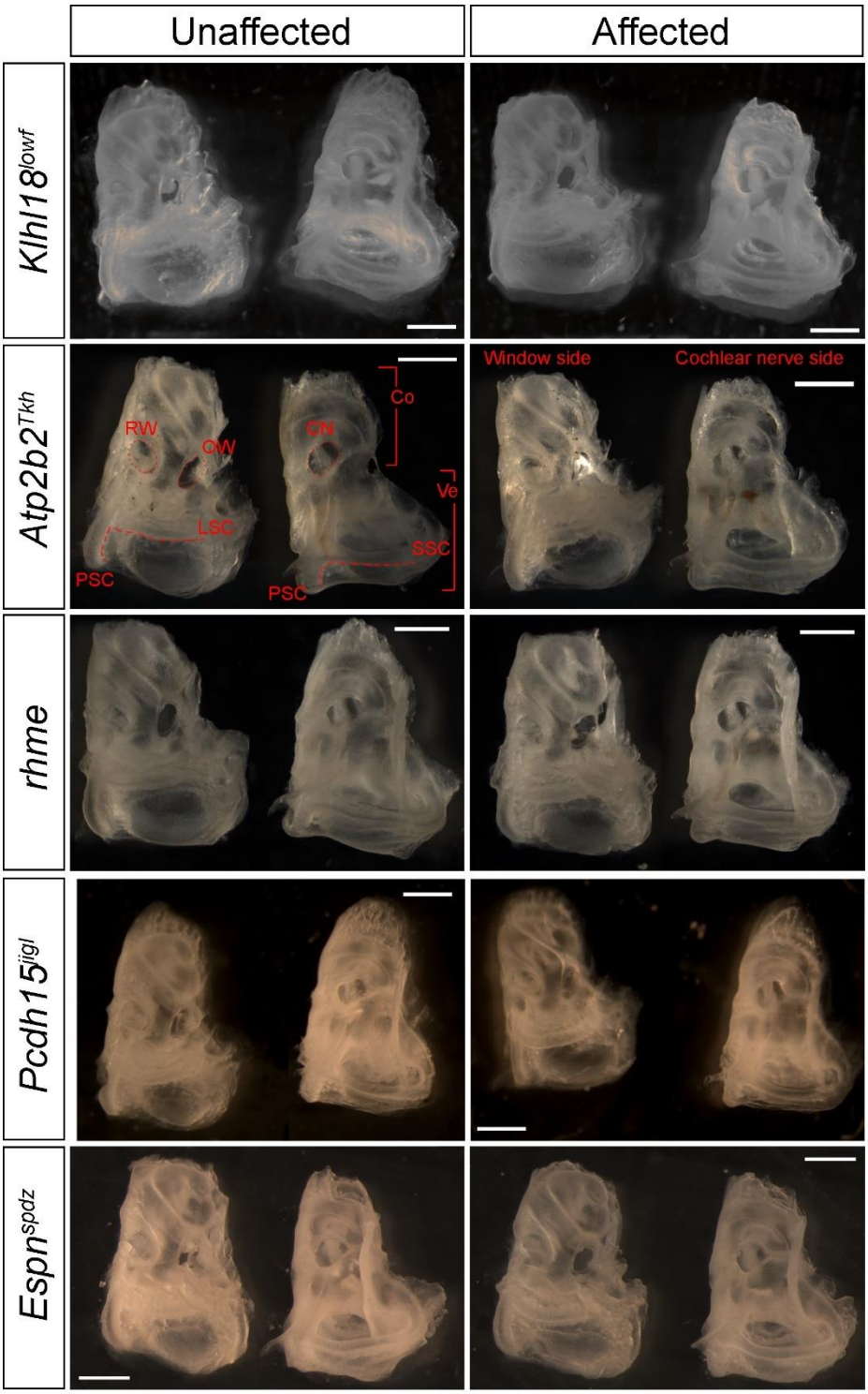
Supplementary Fig. 2. Middle ear ossicles from mice displaying hearing loss show no defects.

Ossicles from affected mice from all seven lines (right) and unaffected littermates (left). Numbers: *Klhl18*^{lowf} n=5 unaffected, 9 affected mice at P94±1day; *Atp2b2*^{Tkh} n=3 unaffected (wildtype), 3 affected (homozygote) mice at P36±2 days; *rhme* n=3 unaffected, 3 affected mice at P69; *Tbx1*^{ttch} n=3 affected, 3 unaffected mice at P28 or older; *Pcdh15*^{jigl} n=7 affected, 4 unaffected mice at P30±1 day; *rthm* n=3 affected, 3 unaffected mice at P28±1 day; *Espn*^{spdz} n=5 affected, 4 unaffected mice at

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

P57 or older. Scale bar = 1mm. For each panel, the malleus is shown on the left, the stapes on the top right and the incus on the bottom right.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



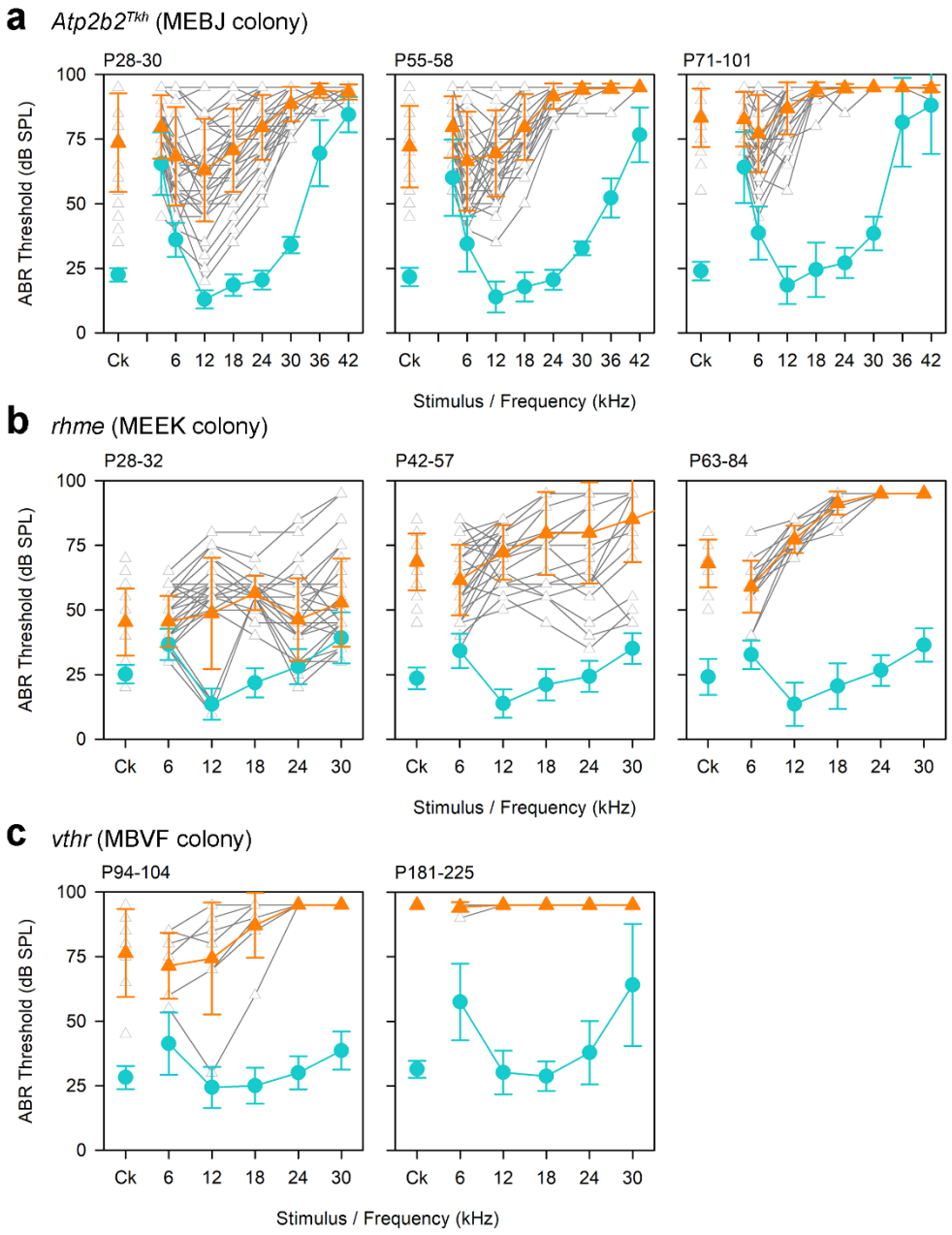
Supplementary Fig. 3. Cleared inner ears from mice displaying hearing loss which show no gross defects.

Cleared inner ears from affected mice carrying the *Klhl18*^{lowf}, *Atp2b2*^{Tkh}, *rhme*, *Pcdh15*^{jigl} and *Espn*^{spdz} alleles (right) and unaffected littermates (left). Numbers: *Klhl18*^{lowf} n=5 unaffected, 7 affected mice at P43±1day; *Atp2b2*^{Tkh} n=3 unaffected (wildtype), 3 affected (homozygote) mice at P36±2 days; *rhme* n=3 unaffected, 3 affected mice at P69; *Pcdh15*^{jigl} n=12 affected, 8 unaffected mice at P30±1

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

day; *Espn*^{spdz} n=5 affected, 4 unaffected mice at P57 or older. Scale bar = 1mm. Labels have been added to the *Atp2b2*^{Tkh} inner ears. The middle ear side, with the round (RW) and oval (OW) windows is shown on the left of each panel, and the brain side, where the cochlear nerve exits (CN), on the right. The round and oval windows and the cochlear nerve exits are marked by dotted lines, and the semicircular canals by dashed lines. Brackets indicate the cochlea (Co) and vestibular region (Ve). LSC=lateral semicircular canal; SSC=superior semicircular canal; PSC=posterior semicircular canal.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



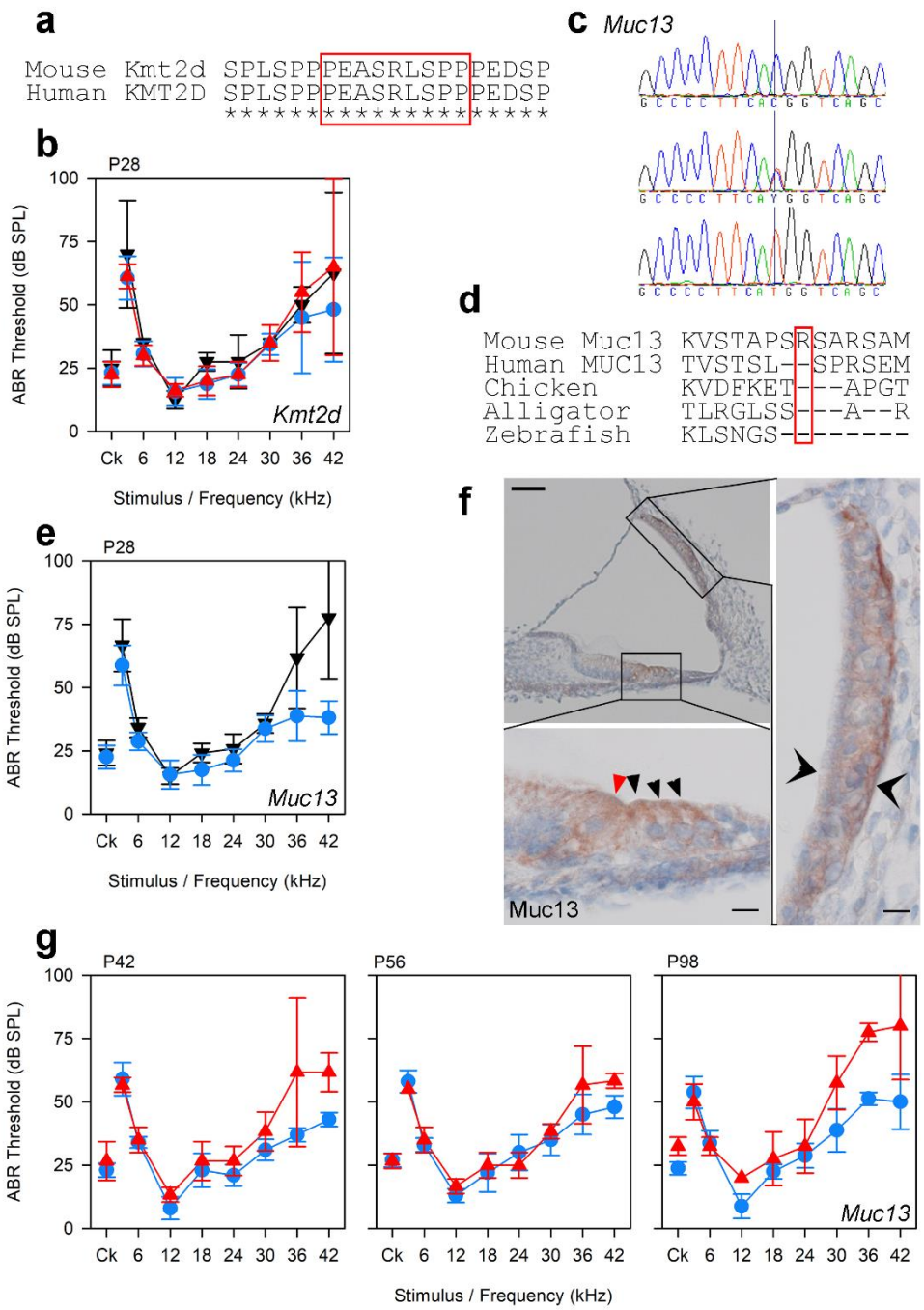
Supplementary Fig. 4. ABR thresholds of mice from the MEBJ, MEEK and MBVF lines showing progressive division into affected and unaffected mice by threshold.

This figure shows ABR thresholds from mice from the MEBJ (a), MEEK (b) and MBVF (c) colonies, showing both the progression of hearing loss with age and the phenotypic division into affected and unaffected mice. These graphs include both genotyped mice (which will also be represented in other figures) and those mice tested before genotyping was possible. Where possible, the same mice were tested at each age. **a** ABR thresholds of affected MEBJ mice carrying the *Atp2b2*^{Tkh} allele and unaffected littermates, showing the progression of hearing loss in affected mice (orange triangles) compared to unaffected mice (teal circles) (n = 10 unaffected and 37 affected at P28-30; n=9 unaffected and 31 affected at P55-58; n=19 unaffected and 37 affected at P71-101). Individual audiograms from affected mice are shown in grey. **b** ABR thresholds from affected (orange triangles) and unaffected (teal circles) *rhme* (MEEK) mice at P28-32 (n = 42 unaffected, 35 affected), P42-57 (n = 83 unaffected, 28 affected) and P63-84 (n = 28 unaffected, 15 affected). Traces from individual affected mice are shown in grey. **c** ABR thresholds of seven MBVF mice carrying the unknown *vthr*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mutation at P94-104 and unaffected mice from the same colony (n=7 affected, 89 unaffected) and P181-225 (n=7 affected, 28 unaffected). The hearing loss appears to be variable at the younger age but progresses to be more severe at ages over 6 months. Traces from individual affected mice are shown in grey.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



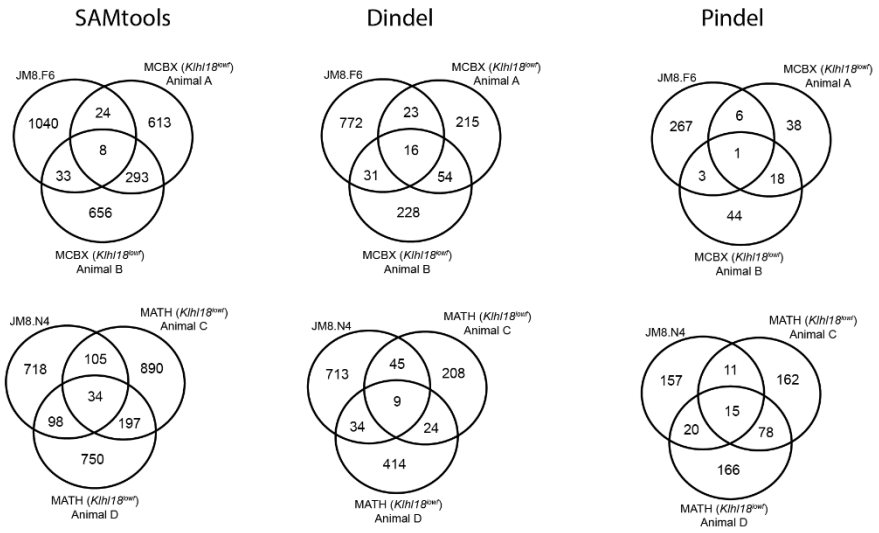
Supplementary Fig. 5. Two other mutations found in the MDLY line carrying the *ttch* allele did not affect hearing.

a Clustal alignment showing deleted amino acids in *Kmt2d* (red box) caused by the g.15:98863687_98863713del mutation, resulting in a loss of 9 amino acids (ENSMUST00000023741, p.586-594del). The region is conserved between mouse and human but not in other vertebrates checked. **b** ABR thresholds for P28 mice homozygous (n=4, red triangles), heterozygous (n=8, blue circles) or wildtype (n=2, black inverted triangles) for the *Kmt2d* deletion. All these mice were heterozygous or wildtype for the *Tbx1*^{ttch} allele, and mice homozygous for the deletion have no difference in their ABR thresholds. **c** Sequence trace showing the missense mutation in *Muc13* (g.16:33807881C>T). **d** Clustal alignment showing the affected amino acid in *Muc13* (red box; p.(Arg334Trp), ENSMUST00000115044). It is not conserved in any other vertebrates checked. **e** ABR

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

thresholds for P28 mice heterozygous (n=8, blue circles) or wildtype (n=6, black inverted triangles) for the *Muc13* missense mutation. All these mice were heterozygous or wildtype for the *Tbx1*^{ttch} allele. **f** MUC13 expression at P4 in a mouse heterozygous for both the *Muc13* missense variant and the *Tbx1*^{ttch} allele (n=2). Brown indicates the presence of MUC13 protein, which is visible in the hair cells (arrowheads; red points to the inner hair cell and black to the outer hair cells) and in the basal cells of the stria vascularis (twin open arrowheads). Scale bar = 50µm in the top left panel and 10µm in the two high magnification panels. The section shown is from the region 50% of the distance along the organ of Corti from base to apex. **g** ABR thresholds for mice homozygous (red triangles) or heterozygous (blue circles) for the *Muc13* mutation at later ages (P42 and P56: n=5 heterozygotes, 3 homozygotes; P98 n=4 heterozygotes, 2 homozygotes). All these mice were heterozygous or wildtype for the *Tbx1*^{ttch} allele.

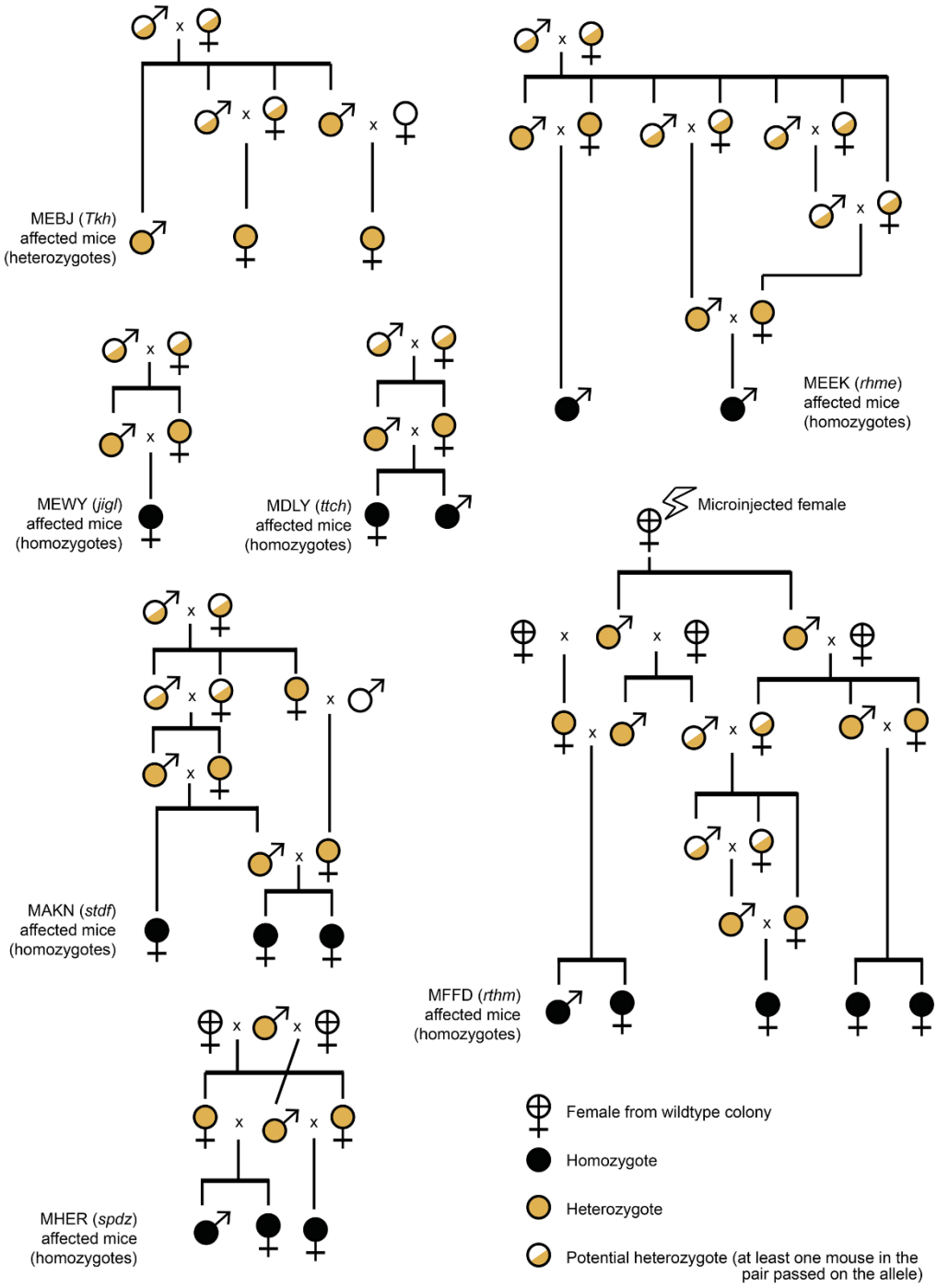
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Fig. 6. A subset of variants identified by each caller were passed on from parental ES cells to mice created using that ES cell line.

Venn diagrams showing the number of high quality variants in two ES cell lines (JM8.F6 and JM8.N4) which were shared by mice from lines created using those ES cell lines (MCBX and MATH), detected by SAMtools, Dindel and Pindel. Both these mouse lines also carry the spontaneous *Klhl18*^{lowf} mutation, which was not seen in any of the ES cell lines. None of the variants identified by BreakDancer were shared between the ES cells and the mice (Supplementary Table 4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Fig. 7. Pedigrees of the affected mice showing the latest possible point at which each mutation could have arisen.

Pedigrees of mouse lines for seven of the eight mutations described here. The *Klhl18*^{lowf} allele is thought to have arisen in the wildtype colony and is not shown. For each of the other lines, the pedigree shows the latest possible point at which the mutation could have arisen. Homozygotes are shown in solid black circles, heterozygotes (obligate carriers for the recessive mutations or mice carrying the semidominant *Atp2b2*^{Tkh} allele) are shown in yellow, and where either or both mice in a pair may have been carrying the mutant allele, this is indicated by the white/yellow colour. In only one line (MFFD, *rthm*) does the mutation definitely originate at or before the point of microinjection

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of the targeted ES cells. Only the mice at the bottom of each pedigree, marked by the text, are known to have been affected (and therefore homozygotes or, for *Atp2b2*^{Tkh}, heterozygotes); all the other genotypes and phenotypes are inferred.

		SAMtools	Dindel	Pindel	Breakdancer
MDLY (*ttch*) Chr16:17663364-20643524	**Initial calls**	14556	8365, 7511	4726	153
	In both affected mice	11773	1049	277	106
	Not in wildtypes	8612	717	164	106
	Not in other sequenced animals (*Tkh* , *rhme* , *lowf*)	5707	367	56	99
	On Chr 16	198	10	1	5
	In region	10	0	0	0
	Exonic	1	0	0	0
MEWY (*jigl*) Chr10:65974036-92894187	**Initial calls**	12111	6651, 7290	4123	155
	In both affected mice	4199	940	257	113
	Not in wildtypes	2816	612	161	113
	Not in other sequenced animals (*Tkh* , *rhme* , *lowf*)	2200	285	46	105
	On Chr 10	219	9	0	5
	In region	57	4	0	2
	Exonic	0	0	0	1 CTX
MFFD (*rthm*) Chr18:56130252-59469675	**Initial calls**	16077	8448, 9239	4919	450
	In both affected mice	12875	1079	278	187
	Not in wildtypes	9528	719	170	187
	Not in other sequenced animals (*Tkh* , *rhme* , *lowf*)	6588	367	52	182
	On Chr18	209	10	0	5
	In region	6	0	0	0
	Exonic	0	0	0	0
MCBX (*lowf*) Chr9:109721879-115398849	**Initial calls**	15592	10245, 9028	5345	203
	In both affected mice	5542	1197	304	131
	Not in wildtypes	3813	813	188	131
	Not in other sequenced animals (*ttch* , *rthm* , *jigl* , *Tkh* , *rhme*)	3032	366	49	122
	On Chr 9	399	24	1	9
	In region	25	2	0	0
	Exonic	3	0	0	0
MEEK (*rhme*) Chr10:15757313-22812579	**Initial calls**	16516	8522, 9112	4856	173
	In both affected mice	13323	1072	299	121
	Not in wildtypes	9834	717	190	121
	Not in other sequenced animals (*ttch* , *rthm* , *jigl* , *lowf*)	6823	323	54	117
	On Chr 10	300	16	1	4
	In region	13	1	0	1
	Exonic	0	0	0	1 CTX
MEBJ (*Tkh*) Chr6:110959922-116155128	**Initial calls**	15411	9270, 8618	4950	162
	In both affected mice	5562	1073	286	122
	Not in wildtypes	3857	726	176	122
	Not in other sequenced animals (*ttch* , *rthm* , *jigl* , *lowf*)	3083	333	55	114
	On Chr 6	159	21	0	7
	In region	6	2	0	1
	Exonic	1	0	0	0
MHER (*spdz*) Chr4:151004833-154007486	**Initial calls**	25552	9047, 13209	320262	22043
	In both affected mice	11725	820	2907	12498
	Not in wildtypes	9632	658	2582	12498
	Not in other sequenced animals (*Tkh* , *rhme* , *lowf*)	9123	571	2437	12498
	On Chr 4	604	35	231	1019
	In region	13	0	0	4
	Exonic	0	0	0	3
MBVF (*vthr*)	**Initial calls**	57323	5865, 4642	2968604	82700
	In both affected mice (het or hom calls)	42100	76	1158	47757
	Not in wildtypes	32142	70	991	47757
	Not in other sequenced animals (*ttch* , *rthm* , *jigl* , *lowf*)	30260	67	854	47757
	After quality filter	2407	41	561	138
	High impact variants	8	0	103	110
MATH (*lowf*)	**Initial calls**	42303	16499, 10811	716806	4476
	In both homs	15618	630	1082	2383
	Not in wildtypes	9141	447	811	2383
	Not in other sequenced animals (*ttch* , *rthm* , *jigl* , *Tkh* , *rhme*)	8117	338	619	2383
	Initial calls	23224	27562	316467	1261

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

JM8.F6	**Not in wildtypes**	14651	22982	2054	1261
	After quality filter	2893	2050	988	53
	High impact variants	9	18	69	31
JM8.N4	**Initial calls**	21903	26427	286051	1197
	Not in wildtypes	13858	21907	1829	1197
	After quality filter	2643	1922	837	53
	High impact variants	17	23	51	32
JM8.N19	**Initial calls**	19242	22780	164035	570
	Not in wildtypes	12297	18951	1881	570
	After quality filter	2631	1759	819	50
	High impact variants	16	16	50	29

Supplementary Table 1. Variant counts from whole exome sequencing.

This table shows the variant counts from each caller for all mice and ES cell line sequenced, and the filtering steps undertaken (the identification of the *S1pr2* ^{stdf} mutation has been previously described[12]). The lefthand column holds the line name, with the allele in brackets where applicable, and the nonrecombinant region identified from mapping, where that was carried out. SAMtools, Pindel and BreakDancer call variants from both animals sequenced, so the single number represents all variants seen in one or both mice. Dindel calls indicate variants per animal. For the "not in wildtype" filter, variants were compared to known variants from the Ensembl database and the Mouse Genomes Project[63,82,83], and only novel variants were kept. We also carried out an additional filtering step, removing any variants seen in the other mice sequenced in this project which had markedly different phenotypes (since the MHER (*spdz*) line was sequenced much later than the others, variants from this line were not used in this filter). For two of the lines, MATH and MBVF (*vthr*), we were not able to map the mutation. In the case of the MATH line, the affected mice displayed the same low frequency hearing loss phenotype as the MCBX line, and we found the same mutation (*Klhl18* ^{lowf}) in all the affected mice we were able to sequence, so we did not filter variants any further. For the ES cells and the *vthr* mice (MBVF colony), we carried out two additional filters; a quality filter (Supplementary Table 6) and an impact filter, where only variants predicted to have a high impact on gene function (by the Ensembl Variant Effect Predictor) passed the filter.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.12.492384 ; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Caller	Location	Variant	Gene	Gene ID	Deafness gene	Consequence
BreakDancer	1:50868574-74855166	deletion	Multiple	Multiple	Yes (Bzw1, Cxcr2, Erbb4, Ikzf2)	
BreakDancer	10:43020918-43023072	deletion	Sobp	ENSMUSG00000038248,	Yes	
BreakDancer	12:110279135-110280943	deletion	Dio3	ENSMUSG00000075707,	Yes	
BreakDancer	14:104465276-104466221	deletion	Pou4f1	ENSMUSG00000048349,	Yes	
BreakDancer	16:30066681-30067674	deletion	Hes1	ENSMUSG00000022528,	Yes	
BreakDancer	17:26067591-26069023	deletion	Rab11fip3	ENSMUSG00000037098,	Yes	
BreakDancer	7:57591027-57591900	deletion	Gabrb3	ENSMUSG00000033676,	Yes	
BreakDancer	X:155622859-155623474	deletion	Ptchd1	ENSMUSG00000041552,	Yes	
Pindel	10:107543351-107543481		Ptprq	ENSMUSG00000035916	Yes	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	6:86950749-86950881		Aak1	ENSMUSG00000057230	Yes	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	13:58392343-91261328	duplication	Multiple	Multiple	Yes (Adgrv1, Ahrr, Med10, Ntrk2, Slc12a7, Tent4a, Zdhhc11)	transcript_amplification
Pindel	4:133638286-147527803	duplication	Multiple	Multiple	Yes (Cdc42, Ephb2, Ifnlr1, Otud3, Tmem51)	transcript_amplification
Pindel	3:56828573-106043216	duplication	Multiple	Multiple	Yes (Chrnb2, Clrn1, Lmna, Lrig2, Wars2)	transcript_amplification
Pindel	4:133638506-147527996	duplication	Multiple	Multiple	Yes (Ephb2, Ifnlr1, Otud3, Tmem51)	transcript_amplification
Pindel	4:133638628-147528105	duplication	Multiple	Multiple	Yes (Ephb2, Ifnlr1, Otud3, Tmem51)	transcript_amplification
BreakDancer	1:127677202-127677707	deletion	Tmem163	ENSMUSG00000026347,		
BreakDancer	1:150453972-150454829	intrachromosomal translocation	Prg4	ENSMUSG00000006014,		
BreakDancer	1:156035702-156036429	deletion	Tor1aip1,Tor1aip2	ENSMUSG00000026466,ENSMUSG00000050565,		
BreakDancer	1:157352211-8:108792045	deletion		Interchromosomal		
BreakDancer	1:171345053-171356260	deletion	Nit1,Pfdn2	ENSMUSG00000013997,ENSMUSG00000006412,		
BreakDancer	1:171356948-171357777	deletion	Pfdn2	ENSMUSG00000006412,		
BreakDancer	1:22533272-22534373	deletion	Rims1	ENSMUSG00000041670,		
BreakDancer	1:28488444-2:143990313	intrachromosomal translocation		Interchromosomal		
BreakDancer	1:36471777-36473095	deletion	Cnnm4	ENSMUSG00000037408,		
BreakDancer	1:3670300-3670919	deletion	Xkr4	ENSMUSG00000051951,		
BreakDancer	1:37471040-11:3189249	inversion		Interchromosomal		
BreakDancer	1:4857823-4857874	deletion	Gm37988,Tcea1	ENSMUSG00000104217,ENSMUSG00000033813,		
BreakDancer	1:49377387-12:49384762	deletion		Interchromosomal		
BreakDancer	1:82539477-82540052	deletion	Agtr1	ENSMUSG00000026159,		
BreakDancer	1:88265031-88267361	deletion	Hjurp	ENSMUSG00000044783,		
BreakDancer	10:123196614-123197231	deletion	Usp15	ENSMUSG00000020124,		
BreakDancer	10:17723768-17723834	deletion	Cited2	ENSMUSG00000039910,		
BreakDancer	10:28074851-28075341	deletion	Ptprk	ENSMUSG00000019889,		
BreakDancer	10:87492756-87492842	deletion	Ascl1	ENSMUSG00000020052,		
BreakDancer	10:9674357-9675080	deletion	Samd5	ENSMUSG00000060487,		
BreakDancer	11:102437618-102438911	deletion	Fam171a2	ENSMUSG00000034685,		
BreakDancer	11:3684168-17:39843063	deletion		Interchromosomal		
BreakDancer	11:6065247-6065792	deletion	Camk2b	ENSMUSG00000057897,		
BreakDancer	11:64895816-7:10382744	inversion		Interchromosomal		
BreakDancer	11:69340645-69340697	deletion	Rnf227	ENSMUSG00000043419,		
BreakDancer	11:81573042-81573097	deletion	Asic2	ENSMUSG00000020704,		
BreakDancer	12:109032788-109034246	deletion	Begain	ENSMUSG00000040867,		
BreakDancer	12:19954256-19954998	deletion	Gm21863	ENSMUSG00000095470,		
BreakDancer	13:11747295-14:19418344	intrachromosomal translocation		Interchromosomal		
BreakDancer	13:28952417-28953074	deletion	Sox4	ENSMUSG00000076431,		
BreakDancer	13:31558834-31560576	deletion	A530084C06Rik,Foxq1	ENSMUSG00000090863,ENSMUSG00000038415,		
BreakDancer	13:31626104-31626119	deletion	Foxf2	ENSMUSG00000038402,		
BreakDancer	13:31658910-18:37307981	deletion		Interchromosomal		
BreakDancer	13:31807247-31807943	deletion	Foxc1	ENSMUSG00000050295,		
BreakDancer	13:34344061-34344733	deletion	Slc22a23	ENSMUSG00000038267,		
BreakDancer	13:69738427-69738964	deletion	Ube2ql1	ENSMUSG00000052981,		
BreakDancer	14:118234744-118234810	deletion	Sox21	ENSMUSG00000061517,		
BreakDancer	14:120478386-120478480	deletion	Rap2a	ENSMUSG00000051615,		
BreakDancer	14:122463634-122464439	deletion	Zic5	ENSMUSG00000041703,		
BreakDancer	14:19415687-19416499	intrachromosomal translocation	Gm21738	ENSMUSG00000095280,		
BreakDancer	14:52196832-52197375	deletion	Supt16	ENSMUSG00000035726,		
BreakDancer	14:54632009-54632074	deletion	Cdh24	ENSMUSG00000059674,		
BreakDancer	14:68123524-68124758	deletion	Nefm	ENSMUSG00000022054,		
BreakDancer	14:79768498-79768501	deletion	Pcdh8	ENSMUSG00000036422,		
BreakDancer	15:27747654-27748626	deletion	Trio	ENSMUSG00000022263,		
BreakDancer	16:22009405-22009940	deletion	Senp2	ENSMUSG00000022855,		
BreakDancer	16:35156077-35157221	deletion	Adcy5	ENSMUSG00000022840,		
BreakDancer	17:13590847-4:156358128	intrachromosomal translocation		Interchromosomal		
BreakDancer	17:34031779-34031839	deletion	Rxrb	ENSMUSG00000039656,		
BreakDancer	17:44734999-44735510	deletion	Runx2	ENSMUSG00000039153,		
BreakDancer	17:5493304-5493959	deletion	Zdhhc14	ENSMUSG00000034265,		
BreakDancer	17:85687199-85688042	deletion	Six2	ENSMUSG00000024134,		
BreakDancer	18:15061756-15063153	deletion	Kctd1	ENSMUSG00000036225,		
BreakDancer	18:37447869-37450801	deletion	Multiple	Multiple		
BreakDancer	19:28964119-28966835	deletion	4430402I18Rik,Plpp6	ENSMUSG00000064202,ENSMUSG00000040105,		
BreakDancer	19:36056385-36057455	deletion	Htr7	ENSMUSG00000024798,		
BreakDancer	19:37173760-37173812	deletion	Cpeb3	ENSMUSG00000039652,		
BreakDancer	19:42147628-42149522	deletion	Marveld1	ENSMUSG00000044345,		
BreakDancer	19:47014920-47014931	deletion	Nt5c2,Ina	ENSMUSG00000025041,ENSMUSG00000034336,		
BreakDancer	19:54046911-54046962	deletion	Adra2a	ENSMUSG00000033717,		
BreakDancer	19:56722375-56723906	deletion	Adrb1	ENSMUSG00000035283,		
BreakDancer	2:131178435-131179908	deletion	Spef1,Cenpb	ENSMUSG00000027329,ENSMUSG00000068267,		
BreakDancer	2:181147847-181148976	deletion	Eef1a2	ENSMUSG00000016349,		

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.12.430925 ; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

BreakDancer	2:26388259-26388314	deletion	Entr1	ENSMUSG00000026927,	
BreakDancer	2:37792984-37793987	deletion	Crb2	ENSMUSG00000035403,	
BreakDancer	2:72979906-72980533	deletion	Sp3,Gm11084	ENSMUSG00000027109,ENSMUSG00000079286,	
BreakDancer	3:10088218-10088659	deletion	Gm9833	ENSMUSG00000049230,	
BreakDancer	3:16183236-16184166	deletion	Ythdf3	ENSMUSG00000047213,	
BreakDancer	3:88568395-X:8073818	deletion		Interchromosomal	
BreakDancer	3:98427007-98428600	deletion	Zfp697	ENSMUSG00000050064,	
BreakDancer	4:120268626-120268683	deletion	Foxo6	ENSMUSG00000052135,	
BreakDancer	4:126735219-126735272	deletion	Tfap2e	ENSMUSG00000042477,	
BreakDancer	4:15266207-15266981	deletion	Tmem64	ENSMUSG00000043252,	
BreakDancer	4:154979319-154980360	deletion	Pank4	ENSMUSG00000029056,	
BreakDancer	4:155221319-155222426	deletion	Ski	ENSMUSG00000029050,	
BreakDancer	4:21678016-21678610	deletion	Prdm13	ENSMUSG00000040478,	
BreakDancer	4:46344988-46346229	deletion	Foxe1	ENSMUSG00000070990,	
BreakDancer	4:99656713-99657509	deletion	Foxd3	ENSMUSG00000067261,	
BreakDancer	5:128601065-128602955	deletion	Fzd10	ENSMUSG00000081683,	
BreakDancer	5:24601545-24602181	deletion	Smarcd3	ENSMUSG00000028949,	
BreakDancer	5:29477428-29478449	deletion	Mnx1	ENSMUSG00000001566,	
BreakDancer	6:113697358-113697413	deletion	Tatdn2	ENSMUSG00000056952,	
BreakDancer	6:116022455-116024673	deletion	Tmcc1	ENSMUSG00000030126,	
BreakDancer	6:120491931-120493384	deletion	Tmem121b	ENSMUSG00000094626,	
BreakDancer	6:30957570-30959193	deletion	Klf14	ENSMUSG00000073209,	
BreakDancer	6:52259661-52259722	deletion	Hoxa13	ENSMUSG00000038203,	
BreakDancer	6:72788241-72789013	deletion	Tcf7l1	ENSMUSG00000055799,	
BreakDancer	6:86524773-86526073	deletion	Pcbp1	ENSMUSG00000051695,	
BreakDancer	6:87980535-87981381	deletion	H1f10	ENSMUSG00000044927,	
BreakDancer	7:142081602-142081662	deletion	Dusp8	ENSMUSG00000037887,	
BreakDancer	7:45406016-45406077	deletion	Kcna7	ENSMUSG00000038201,	
BreakDancer	8:11556710-11557302	deletion	Ing1	ENSMUSG00000045969,	
BreakDancer	8:122335434-122335495	deletion	Galnt2l,Zfpm1	ENSMUSG00000092329,ENSMUSG00000049577,	
BreakDancer	8:124895469-124897710	deletion	Exoc3	ENSMUSG00000074030,	
BreakDancer	8:45225221-45227292	deletion	Tenm3	ENSMUSG00000031561,	
BreakDancer	8:70838258-70839570	deletion	Arrdc2	ENSMUSG00000002910,	
BreakDancer	8:85690424-85690993	inversion	Neto2	ENSMUSG00000036902,	
BreakDancer	9:106884376-106886227	deletion	Manf,Rbm15b	ENSMUSG00000032575,ENSMUSG00000074102,	
BreakDancer	9:108692265-108692361	deletion	Prkar2a	ENSMUSG00000032601,	
BreakDancer	9:3025295-3025606	intrachromosomal translocation	Gm10717	ENSMUSG00000095891,	
BreakDancer	9:31280744-31280847	deletion	Prdm10	ENSMUSG00000042496,	
BreakDancer	9:95560271-95560322	deletion	Paqr9	ENSMUSG00000064225,	
Pindel	1:171070969-171078593		Gm27552	ENSMUSG00000098356	transcript_amplification
Pindel	1:171070973-171078590		Gm27552	ENSMUSG00000098356	transcript_ablation
Pindel	1:171345008-171356527		Pfdn2,Nit1	ENSMUSG00000006412,ENSMUSG00000013997	splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant,splice_acceptor_variant
Pindel	1:171356617-171356702		Pfdn2	ENSMUSG00000006412	splice_donor_variant,coding_sequence_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	1:33777803-33777917		Zfp451	ENSMUSG00000042197	splice_acceptor_variant,coding_sequence_variant,intron_variant,3_prime_UTR_variant,NMD_transcript_variant
Pindel	1:85099381-85259607		C130026I21Rik,Gm2619	ENSMUSG00000052477,ENSMUSG00000091199	splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	1:87989996-87990135		Usp40	ENSMUSG00000005501	splice_donor_variant,coding_sequence_variant,intron_variant,non_coding_transcript_exon_variant,3_prime_UTR_variant,NMD_transcript_variant
Pindel	1:88232258-88232259	CC/-	Mroh2a	ENSMUSG00000079429	frameshift_variant
Pindel	1:88256166-88256167	-/T	Mroh2a	ENSMUSG00000079429	frameshift_variant
Pindel	1:88266524-88266525	-/T	Hjurp	ENSMUSG00000044783	frameshift_variant,stop_gained
Pindel	10:58221207-58233780		AW822073,Duxf3,Gm10807	ENSMUSG00000058537,ENSMUSG00000075046,ENSMUSG00000057216	transcript_amplification
Pindel	10:58224025-58231154		Duxf3	ENSMUSG00000075046	splice_acceptor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	10:58231158-58231706		Duxf3	ENSMUSG00000075046	splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	10:58233778-58237853		Gm4981	ENSMUSG00000075045	transcript_amplification
Pindel	10:58233811-58237886		Gm4981	ENSMUSG00000075045	transcript_ablation
Pindel	10:58233944-58238022		Gm4981	ENSMUSG00000075045	transcript_amplification
Pindel	10:58233945-58238019		Gm4981	ENSMUSG00000075045	transcript_ablation
Pindel	11:33156083-33156084	-/T	Npm1	ENSMUSG00000057113	frameshift_variant
Pindel	12:55863034-55863165		Brms1l	ENSMUSG00000012076	splice_acceptor_variant,coding_sequence_variant,intron_variant
Pindel	13:100227562-100301267		Naip6,Naip5,Naip3-ps1	ENSMUSG00000078942,ENSMUSG00000071203,ENSMUSG00000114576	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant,5_prime_UTR_variant,non_coding_transcript_exon_variant
Pindel	13:104312805-104312913		Adamts6	ENSMUSG00000046169	stop_gained,NMD_transcript_variant
Pindel	14:19415782-19418819		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	14:19415817-19417804		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	14:19415817-19419322		Gm21738	ENSMUSG00000095280	transcript_amplification
Pindel	14:19415875-19417862		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19415995-19419617		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant
Pindel	14:19416241-19418693		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416245-19418697		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416246-19418698		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416249-19418701		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,intron_variant

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.29.450362; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Pindel	14:19416256-19418708		Gm21738	ENSMUSG00000095280	splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416259-19418711		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416268-19418720		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416395-19418496		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19416938-19418571		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19417004-19418754		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19417006-19418756		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19417054-19418571		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	14:19417171-19418571		Gm21738	ENSMUSG00000095280	splice_acceptor_variant,coding_sequence_variant,intron_variant
Pindel	14:33624874-33624984		Ptpn20	ENSMUSG00000021940	splice_acceptor_variant,coding_sequence_variant,intron_variant,3_prime_UTR_variant,NMD_transcript_variant
Pindel	14:3545914-4969171	deletion	Multiple	Multiple	transcript_ablation,stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation
Pindel	14:41518121-42219186	deletion	Multiple	Multiple	transcript_ablation
Pindel	15:58111116-58111290		Atad2	ENSMUSG00000022360	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	17:23345092-23385951		Vmn2r116,Vmn2r115	ENSMUSG00000090966,ENSMUSG00000091076	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant,3_prime_UTR_variant
Pindel	17:23446140-23471274		Vmn2r117	ENSMUSG00000091407	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	2:103757666-103757798		Nat10	ENSMUSG00000027185	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	2:15616172-15616643		Rhbdf2	ENSMUSG00000027620	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant,NMD_transcript_variant,3_prime_UTR_variant
Pindel	2:158723946-158723953		Pip4k2a	ENSMUSG00000026737	splice_acceptor_variant,coding_sequence_variant,intron_variant
Pindel	2:98662758-98666465		Gm10801	ENSMUSG00000075015	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	2:98662783-98666250		Gm10801	ENSMUSG00000075015	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	2:98662807-98666274		Gm10801	ENSMUSG00000075015	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	2:98662807-98666395		Gm10801	ENSMUSG00000075015	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	2:98663804-98663805	-/GTC	Gm10801	ENSMUSG00000075015	splice_acceptor_variant
Pindel	3:55994010-55994116		Nbea	ENSMUSG00000027799	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant
Pindel	4:145391472-145574972		Multiple	Multiple	transcript_amplification
Pindel	4:145391514-145575014		Multiple	Multiple	transcript_ablation
Pindel	4:145391559-145575059		Multiple	Multiple	transcript_amplification
Pindel	4:145546367-147343374	duplication	Multiple	Multiple	transcript_amplification
Pindel	4:145585060-147449053	duplication	Multiple	Multiple	transcript_amplification
Pindel	4:145585074-147449066	duplication	Multiple	Multiple	transcript_amplification
Pindel	4:145774684-147125220	deletion	Multiple	Multiple	transcript_ablation
Pindel	4:146240667-146581346	deletion	Multiple	Multiple	transcript_ablation
Pindel	4:146974010-147807118		Multiple	Multiple	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,3_prime_UTR_variant,intron_variant,non_coding_transcript_exon_variant
Pindel	4:156333858-156350560		Vmn2r125	ENSMUSG00000096042	splice_acceptor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant
Pindel	4:156343290-156357522	deletion	Vmn2r125	ENSMUSG00000096042	stop_lost,coding_sequence_variant,5_prime_UTR_variant,3_prime_UTR_variant,intron_variant
Pindel	4:60662424-61439751	duplication	Multiple	Multiple	transcript_amplification
Pindel	5:145631608-145862324		Cyp3a11	ENSMUSG00000056035	splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	5:145702064-145862281		Cyp3a11	ENSMUSG00000056035	splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Pindel	5:14952638-15007307		Gm10354	ENSMUSG00000072188	transcript_ablation
Pindel	5:14986071-15040183		Gm17019	ENSMUSG00000091897	transcript_amplification
Pindel	5:15007309-15508974	duplication	Gm17019,Gm43391,Gm21149	ENSMUSG00000091897,ENSMUSG00000106004,ENSMUSG00000099762	transcript_amplification
Pindel	5:15007312-15508977	deletion	Gm17019,Gm43391,Gm21149	ENSMUSG00000091897,ENSMUSG00000106004,ENSMUSG00000099762	transcript_ablation
Pindel	5:15040761-15549297	deletion	Gm21190,Gm21149	ENSMUSG00000106445,ENSMUSG00000099762	stop_lost,coding_sequence_variant,5_prime_UTR_variant,3_prime_UTR_variant,intron_variant
Pindel	5:15463309-15522207		Gm21847,Gm43391,Gm21149	ENSMUSG00000094230,ENSMUSG00000106004,ENSMUSG00000099762	splice_donor_variant,non_coding_transcript_exon_variant,intron_variant,splice_acceptor_variant,coding_sequence_variant,5_prime_UTR_variant,3_prime_UTR_variant
Pindel	5:26121754-27501771	duplication	Multiple	Multiple	transcript_amplification
Pindel	6:18220585-18220695		Cftr	ENSMUSG00000041301	splice_acceptor_variant,coding_sequence_variant,intron_variant,NMD_transcript_variant
Pindel	6:52179268	C/-	5730596B20Rik	ENSMUSG00000056468	frameshift_variant
Pindel	7:24097894-24101445		Zfp180	ENSMUSG00000057101	splice_donor_variant,coding_sequence_variant,intron_variant,3_prime_UTR_variant,NMD_transcript_variant,non_coding_transcript_exon_variant
Pindel	7:24101568-24104403		Zfp180	ENSMUSG00000057101	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,intron_variant,3_prime_UTR_variant,NMD_transcript_variant
Pindel	8:26160857-26160869	AGCATCTGCTCGG/-	Thap1	ENSMUSG00000037214	frameshift_variant

Pindel	8:54681445-54681559		Wdr17	ENSMUSG00000039375	splice_acceptor_variant,non_coding_transcript_exon_variant,intron_variant,coding_sequence_variant
Pindel	8:85302773-85302873		Orc6	ENSMUSG00000031697	splice_acceptor_variant,coding_sequence_variant,intron_variant,5_prime_UTR_variant,NMD_transcript_variant
Pindel	9:103355190-103355204	CTGATTCTCAAGCAC/TA	Cdv3	ENSMUSG00000032803	frameshift_variant
Pindel	9:123873050-124257013		Multiple	Multiple	transcript_ablation
Pindel	9:123873053-124257109	deletion	Multiple	Multiple	transcript_ablation
Pindel	9:3003766	T/-	Gm11168	ENSMUSG00000096385	frameshift_variant
Pindel	9:3016717	T/-	Gm10720	ENSMUSG00000074564	frameshift_variant
Pindel	9:3025485-3025486	-/T	Gm10717	ENSMUSG00000095891	frameshift_variant
Pindel	9:3028641	C/-	Gm10717	ENSMUSG00000095891	frameshift_variant
Pindel	9:3030703-3030818		Gm10717	ENSMUSG00000095891	frameshift_variant
Pindel	9:3030802-3030919		Gm10717	ENSMUSG00000095891	splice_acceptor_variant,coding_sequence_variant,intron_variant,frameshift_variant
Pindel	GL456210.1:20627-167653		Multiple	Multiple	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,3_prime_UTR_variant,intron_variant
Pindel	GL456210.1:33958-163120		Multiple	Multiple	splice_donor_variant,5_prime_UTR_variant,intron_variant,splice_acceptor_variant,coding_sequence_variant,3_prime_UTR_variant
Pindel	GL456210.1:34012-163066		Multiple	Multiple	splice_donor_variant,5_prime_UTR_variant,intron_variant,splice_acceptor_variant,coding_sequence_variant,3_prime_UTR_variant
Pindel	GL456221.1:24738-82426		AC132444.1,AC132444.3,AC132444.5	ENSMUSG00000079222,ENSMUSG00000094874,ENSMUSG00000095500	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant,intron_variant,3_prime_UTR_variant
Pindel	GL456221.1:36961-64957		AC132444.5	ENSMUSG00000095500	splice_acceptor_variant,splice_donor_variant,coding_sequence_variant,3_prime_UTR_variant,intron_variant
Samtools	1:191354614-191354618	AAGAA/-	Ppp2r5a	ENSMUSG00000026626	frameshift_variant
Samtools	13:107889579	C/CC	Zswim6	ENSMUSG00000032846	splice_acceptor_variant,non_coding_transcript_exon_variant
Samtools	15:98672500-98672501	TG/-	Rnd1	ENSMUSG00000054855	frameshift_variant
Samtools	17:23311107-23311109	CCC/CC	Vmn2r114	ENSMUSG00000091945	frameshift_variant
Samtools	17:23311191-23311195	AAAAA/AAAA	Vmn2r114	ENSMUSG00000091945	frameshift_variant
Samtools	2:131274223-131274224	-/AT	Pank2	ENSMUSG00000037514	frameshift_variant
Samtools	2:24762959-24762961	CCG/CCCG	Cacna1b	ENSMUSG00000004113	frameshift_variant
Samtools	2:43899815-43899816	-/C	Acvr2a	ENSMUSG00000052155	frameshift_variant

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.31.437922 This version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Table 2. High quality, high impact variants identified in *vthr* mice (MBVF colony)

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-4.0 International license.

a Genotyping and mutation identification primers

Primer name	Primer sequence	Reverse primer name	Reverse primer sequence	Comments
Lowf_F	GAAGTGTAAGGGTGTGGGGG	Lowf_R	CTTGCCCGTCTGTCATCCC	Mutation can be identified by sequencing
Tkh_F	TCCCAGGCCTATGAAGATGC	Tkh_R	ACTAAGCAGGAACCACAGGC	Mutation can be identified by sequencing
Ttch_F	TGCCCCTGTCTCTAGTCTGG	Ttch_R	TCCACAGTCCCCTCTCTTGG	Mutation can be identified by sequencing
Rthm_F	CCTCCTTAGTGCTGGGGGAACAG	Rthm_mutR	ACTGAGTCCTTGTTGAATTCATTT	250bp in mutant allele
Rthm_F	CCTCCTTAGTGCTGGGGGAACAG	Rthm_wtR	AATGAGTGAGGAAGGACGGC	645bp in wildtype allele
Rhme_F	GAGGGTCAAAGGTTCGATCC	Rhme_mutR	CCATGGAAAGTACACATTCTTGG	350bp in mutant allele
Rhme_F	GAGGGTCAAAGGTTCGATCC	Rhme_wtR	CATATGTGCAGGCCCAAAGG	225bp in wildtype allele
Spdz_testF	ACTATCCCTCTAGCCCCACG	Spdz_testR	AGCTGGAGTGGTTATGAGCC	No band is obtained from mutant allele; 688bp band in wildtype allele
Spdz_controlF	GGTCTTGACGTATTGCAGGC	Spdz_controlR	CCAGTGTATGAAAACTATTATTTCAG	Control primers (659bp band should be obtained from mutant and wildtype alleles)
Jigl_1F	ACTCTGCGCCCAAGGTAC	Jigl_1R	ACACCCAAAGAAAATACTAGGCA	No band is obtained from mutant allele. Primers used for testing parental ES cell line and P5 pups collected for SEM
Jigl_controlF	GTGAGACACCAAAGAATACCT	Jigl_controlR	CCCCATGCATTCACATACACT	Control primers (band should be obtained from mutant and wildtype alleles). Primers used for testing parental ES cell line and P5 pups collected for SEM
Jigl_2F	GGGTGTCCTATTCGTCCACC	Jigl_2R	GAGTATTTACCTGGCTCGTGG	No band is obtained from mutant allele. Primers used for screening colony
Muc13_F	AGCACACACAGCAGACTTGG	Muc13_R	CTACACACACACAGAGGGGG	Mutation can be identified by sequencing
Kmt2d_F	ACAGATGTGACAGGACAGGC	Kmt2d_R	CGAAGAGTCACCCATGTCCC	Mutation can be identified by sequencing
b Splicing primers				
Espin_1F	CCACAGGCTACCTCTCTTGC			
Espin_2F	TGCCTATCCACTACGCTGC			
Espin_5F	CGACCTGGCAGAGTTCAATG			
Espin_6F2	CAATGACCTTTGACCTGGGC			
Espin_6R	CAAGACTCGGTGTTCCAAGC			
Espin_7F	AGACCAAGAACAAGCTTCGC			
Espin_7R	AAGCTTGGTGGTGGTGGT			
Espin_9F	CAAACAGCGCTCGTTCAGTA			
Espin_10F	GTTCCTCGTCTTCTACTGGC			
Espin_10R2	TTTGGAGTCCGCTGAGCAG			
Espin_13R	GTGACACCAGAGGCTGCG			
Espin_14R	CTCCTCCTCCAGCTTCTTCC			
Espin_16R	CCCTCTGTATTTCCTCCAGC			
Espin_12-13R	CTGCGGTGACTCAGGCTG			
Espin_30446R	ccgggagaagcagtaggg			
Espin_12-13F	CTCCCAGCCTGAGTCACC			
Espin_30446F	ccctactgcttctcccgg			

Supplementary Table 3. Primer sequences for primers designed in this study.
a Primers used for genotyping spontaneous alleles. **b** Primers used to examine *Espn* mRNA splicing in mice carrying the *spdz* allele.

a

		SAMtools	Dindel	Pindel	Breakdancer
JM8.F6	Initial calls	23224	27562	316467	1261
	Not in wildtypes	14651	22982	2054	1261
	Not in other ES cells	8611	20094	963	1253
	High quality variants	1105	842	277	47
	High impact variants	3	9	25	26
LOWF12.1c	Initial calls	8466	10245	740	181
	Not in wildtypes	5355	8357	546	181
	Not in other ES cells	3772	7241	284	181
	High quality variants	938	308	63	5
	High impact variants	23	12	13	3
LOWF68.2b	Initial calls	8112	9028	685	153
	Not in wildtypes	5271	7423	507	153
	Not in other ES cells	3801	6353	261	153
	High quality variants	990	329	66	5
	High impact variants	32	21	9	3
JM8.N4	Initial calls	21903	26427	286051	1197
	Not in wildtypes	13858	21907	1829	1197
	Not in other ES cells	7763	19076	759	1190
	High quality variants	955	801	203	49
	High impact variants	7	12	14	31
MATH38.1f	Initial calls	21356	10811	2298	3568
	Not in wildtypes	12922	9413	1862	3568
	Not in other ES cells	8947	8571	973	3562
	High quality variants	1226	286	266	19
	High impact variants	10	3		10
MATH38.1g	Initial calls	22260	16499	2289	3291
	Not in wildtypes	14094	13994	1850	3291
	Not in other ES cells	8575	12563	889	3285
	High quality variants	1079	481	279	23
	High impact variants	7	10	40	16

b

Predicted mutation	ES cells and mice	Caller	Confirmed	Primer F	Primer R
1:71642993A>C	JM8.F6 and MCBX mice	SAMtools	Yes; the ES cell and one of the mice are heterozygous for this variant, the other mouse is a homozygote	CAGCATTCAGGAAGCAGAGG	TTACCAACCGCAGACTCACC
11:3188470AT>A	JM8.F6 and MCBX mice	Dindel	There is a homozygous variant present in ES cell and both mice but it is ATCTT>AAAA	TAGGCTGTGTAGGAGGGTGG	GTGTAGCGTTCATTCCAGTGC
7:13803939T>TTA	JM8.F6 and MCBX mice	Dindel	No	GGTTCCTCCCTTCTGCTTCC	GACACTCCACTTCTCCCTGC
9:24542077CA>C	JM8.F6 and MCBX mice	Dindel	No	CCAGAGTGAAGAGGCAGGC	CCTCTGAGTTGTTAGCCAGGG
5:15033229TCTAA>T	JM8.F6 and MCBX mice	Dindel	No	TACAGCCTTTGGTTCCCTGC	CAAGGCTTCATGAGTGTGGC
5:137300304A>C	JM8.F6 and MCBX mice	SAMtools	No	GTAACCTCCCAGCCAAGACC	CTTCTGTGGGTGTGAGAGGC

7:42613972G>A	JM8.F6 and MCBX mice	SAMtools	No	TGTGGATAGGCTTTATTACACTGC	TGCAGGCACGAAAGAAGTCC
7:44998640T>G	JM8.F6 and MCBX mice	SAMtools	No	GTTTTCTGCAAGGGTGAGGC	CTTCCACGGGATCCTGAACC
15:78760942A>C	JM8.F6 and MCBX mice	SAMtools	No	CATAGCCCCTCTCATCCACC	AAGTCCAGGCTTCAGTGACG
10:120928031C>A	JM8.F6 and MCBX mice	SAMtools	No	ACATAACCGGGTCAAACATATTTCG	GGCCAAAACCACCTTCTTCC
11:51651118C>CG	JM8.F6 and MCBX mice	Dindel	Failed	AAAACCCGGCCTGACTCACC	GAGACCTGACCAACCCATCC
16:97850976TG>T	JM8.N4 and MATH mice	Pindel	No	CATGGGAGGATTCTGACGGG	CTGGGAAAAGGGTGGACAGG
1:88232257GCC>G	JM8.N4 and MATH mice	Dindel	No	GTAGCTCCTCGATTCCCTGC	ACGCACCACTACAGATCACC
2:41988800C>T	JM8.N4 and MATH mice	SAMtools	No	CTTCTTCTCAAACTCGCCCG	CAGTCAGAGGGTGGATTGGG
2:50911333G>T	JM8.N4 and MATH mice	SAMtools	No	TGGAGTGTGTGTGGAAACGG	AGTGTTTCATGTTGGCTGTCC
2:164997480A>C	JM8.N4 and MATH mice	SAMtools	No	AGTTTTGCTCAACATGCCCG	TCGCCACCTTTATTGCAACG
1:88248839G>T	JM8.N4 and MATH mice	SAMtools	No	GTGTCTGTGTGAGTCCAGGG	TGTGTGGTAGTCTGCATTGGG
12:88130930C>T	JM8.N4 and MATH mice	SAMtools	No	TGTCCCCTATGTTTCCACGC	TCAAAACAGATCTGCGCTTCC
7:13403237G>T	JM8.N4 and MATH mice	SAMtools	No	AACCTGAGAACTTGGCTGGG	TGGACACCAATCATCCCTGC
7:38430035C>CT	JM8.N4 and MATH mice	Pindel	Failed	CATGGCATTGACAGCTCAGC	GCAGAAACAGAGCAGTCACC
2:41988731G>A	JM8.N4 and MATH mice	SAMtools	Failed	GTTTTCTGCAAGGGTGAGGC	CTTCCACGGGATCCTGAACC

Supplementary Table 4. ES cell variant counts and confirmation.

a This table shows the variant counts from each caller for individual mice and their ancestral ES cells, and the filtering steps undertaken. The process is similar to that described for Supplementary Table 1, but here instead of filtering by variants found in other mouse lines, we eliminated variants found in other ES cells. The quality filter steps are defined in Supplementary Table 6, and high impact variants were those predicted by the Variant Effect Predictor to have a high impact on gene function. **b** Variant validation using Sanger sequencing and primers used. Most variants were not validated. Three of them failed (either by failure to amplify or failure to obtain good quality sequencing), and two were confirmed in the JM8.F6 ES cell line and its descendant mice, although one of them, the small indel called by Dindel, was misidentified.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Chr12	Chr13	Chr14	Chr15	Chr16	Chr17	Chr18	Chr19
D1mit159	D2MIT128	D3mit19	D4mit133	D5Jcs313	D6Mit104	D7Mit31	D8Mit124	D9mit12	D10Mit106	D11mit140	D12Mit106	D13Mit151	D14MIT152	D15mit175	D16Mit12	D17Mit113	D18Mit120	D19mit13
D1mit173	D2MIT156	D3mit199	D4Mit15	D5Mit13	D6Mit150	D7Mit340	D8Mit280	D9mit18	D10Mit12	D11mit229	D12Mit158	D13Mit210	D14MIT225	D15mit209	D16Mit152	D17mit123	D18mit184	D19mit137
D1mit209	D2mit200	D3mit203	D4Mit166	D5Mit157	D6mit159	D7Mit350	D8Mit292	D9mit212	D10Mit168	D11mit270	D12Nds2	D13Mit88	D14MIT234	D15mit245	D16mit165	D17Mit125	D18mit222	D19Mit16
D1MIT353	D2mit241	D3MIT258	D4Mit172	D5Mit277	D6MIT254	D7Mit66	D8Mit45	D9mit214	D10Mit188	D11mit301		D13Mit9	D14Mit266		D16Mit4	D17mit160	D18Mit35	
D1mit380	D2Mit369	D3mit321	D4Mit203	D5Mit292	D6mit268	D7Mit81		D9Mit224	D10Mit189	D11mit62			D14Mit99		D16mit63	D17MIT218	D18mit48	
D1mit415	D2mit493	D3mit339	D4Mit284	D5Mit345	D6mit29			D9mit269	D10Mit194	D11mit71					D16mit86	D17mit28	D18Mit7	
D1mit429	D2mit94		D4Mit33	D5Mit346	D6mit366			D9mit300	D10Mit206	D11mit99					rs4164761		rs3707088	
D1Mit445			D4Mit71	D5Mit391	D6mit59			D9mit305	D10Mit212						rs4165069		rs3681210	
D1mit495			rs31956009		D6mit93			D9mit346	D10Mit213						rs4165081		rs29822437	
			rs32715396					D9mit355	D10Mit214						rs4165103		rs30259125	
			rs32728225					D9mit50	D10Mit247						rs4165193		rs3682219	
			rs31870527						D10mit271						rs4165279		rs30164494	
			rs32830963						D10Mit3						rs4165287		rs3715746	
			rs31795576						D10mit42						rs4165440		rs29630896	
			rs32489175						D10mit68						rs4166223		rs3720827	
			rs32928505						D10Mit80						rs4167316			
			rs32333709						rs6314360									
			rs32553370						rs3725109									
			rs32259888						rs29334568									
			rs31898627						rs29340940									
			rs32901419						rs29346321									
			rs33891699						rs29334604									
									rs224139634									
									rs29314730									
									rs29375570									
									rs29337727									

Supplementary Table 5. Markers used for genome scans and fine mapping.

a

Step	Software	Reference
Fastq quality checks (Spdz only)	Trimmomatic	Bolger et al, 2014[61]
Fastq trimming (Spdz only)	FastQC	Andrews et al, 2010[60]
Mapping to reference sequence	bwa	Li et al, 2009[59]
Mapping to reference sequence (Spdz only)	hisat2	Kim et al, 2015[62]
Local realignment around insertions and deletions	GATK	McKenna et al, 2010[64]
Mark Duplicates, Collect Multiple Metrics	Picard	Broad Institute[65]
Sam/bam file handling, incl. index, fixmate	Samtools	Li et al, 2009[59]
SNV calling	Samtools	Li et al, 2009[59]
Small indel calling	Dindel	Albers et al, 2011[67]
Large indel calling	Pindel	Ye et al, 2009[69]
Structural variant calling	Breakdancer	Chen et al, 2009[68]
Variant annotation	Ensembl Variant Effect Predictor (VEP)	McLaren et al, 2016[70]

b

Caller	Minimum read depth	Other score	Filter mini	Impact filter
Samtools	10	Mapping quality	45	VEP prediction of high impact
Dindel	10	Filter	PASS	VEP prediction of high impact
Pindel	20	Filter	PASS	VEP prediction of high impact
BreakDancer	20	Score	50	Breakpoints cover one or more genes or suggest interchromosomal rearrangements

c

Sample	ENA Accession number
MDLY mouse 1	ERS383294
MDLY mouse 2	ERS383295
MEWY mouse 1	ERS383296
MEWY mouse 2	ERS383297
MFFD mouse 1	ERS383298
MFFD mouse 2	ERS383299
MCBX mouse 1	ERS383300
MCBX mouse 2	ERS383301
MEBJ mouse 1	ERS383302
MEBJ mouse 2	ERS383303
MEEK mouse 1	ERS383304
MEEK mouse 2	ERS383305
MAKN mouse 1	ERS039389
MAKN mouse 2	ERS039388
MATH mouse 1	ERS039394
MATH mouse 2	ERS039395
MBVF mouse 1	ERS041154
MBVF mouse 2	ERS041155
MHER mouse 1	ERS6599514
MHER mouse 2	ERS6599515
JM8.F6	ERS040145
JM8.N4	ERS040146
JM8.N19	ERS040147

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.30.450312; this version posted June 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Table 6. Sequence processing and filtering details.

a Software used for processing and aligning reads, and calling and annotating variants. **b** Thresholds chosen for the variant quality filtering carried out for the MBVF line and ES cell whole exome sequencing. **c** ENA accession numbers for individual samples.