

Rapid Identification of a Disease Allele in Mouse Through Whole Genome Sequencing and Bulk Segregation Analysis

Carrie N. Arnold,* Yu Xia,* Pei Lin,* Charles Ross,* Martin Schwander,[†] Nora G. Smart,*
Ulrich Müller[†] and Bruce Beutler*^{*,1}

*Department of Genetics and [†]Department of Cell Biology, Institute for Childhood and Neglected Disease,
The Scripps Research Institute, La Jolla, California 92037

Manuscript received October 25, 2010
Accepted for publication December 29, 2010

ABSTRACT

In a pedigree of C57BL/6J mice homozygous for germline mutations induced by the mutagen *N*-ethyl-*N*-nitrosourea (ENU), numerous animals died under specific pathogen-free (SPF) conditions between 6 and 7 months of age. Death was caused by nephritic syndrome, which progressed to renal failure associated with focal segmental glomerulosclerosis. To identify the mutation responsible for renal disease, we sequenced genomic DNA from an affected animal using the Applied Biosystems SOLiD sequencing platform. Approximately 74% of the nucleotides comprising coding sequences and splice junctions in the mouse genome were covered at least three times. Within this portion of the genome, 64 discrepancies were flagged as potential homozygous mutations and 82 were flagged as potential heterozygous mutations. A total of 10 of these calls, all homozygous, were validated by capillary sequencing. One of the validated mutations disrupted splicing of the *Col4a4* transcript. Genetic mapping by bulk segregation analysis excluded all mutations but this one as the cause of renal disease in *Aoba* mice. *Col4a4* has not been targeted in the mouse, and this strain, named *Aoba*, represents the first functionally null allele in this species. Our study demonstrates the speed and utility of whole genome sequencing coupled with low resolution meiotic mapping as a means of identifying causative mutations induced by ENU.

THE laboratory mouse has been used as a model organism in forward genetic studies in which the point mutagen, *N*-ethyl-*N*-nitrosourea (ENU), is applied to generate phenotype, and positional cloning is used to find causative mutations. The availability of an annotated reference genome sequence for the C57BL/6J strain has greatly accelerated the identification of point mutations. However, it remains necessary to outcross mutations and formally map them. In the past, “fine mapping” was often needed to define a critical region that could be directly sequenced at the genomic level to identify causative mutations.

Coarse mapping to a large chromosomal region coupled with whole genome sequencing using massively parallel short-read sequencing platforms may be sufficient to exclude all but a single ENU-induced mutation in a defined genetic background. Whole genome sequencing may thereby eliminate the need for fine mapping of mutations and exploration of large critical regions, both expensive and time-consuming steps in positional cloning. Whereas a year or more was formerly required to identify causative mutations, a month or less may now be needed.

Supporting information is available online at <http://www.genetics.org/cgi/content/full/genetics.110.124586/DC1>.

¹Corresponding author: 10550 N. Torrey Pines Rd., Mail Code SP-293, La Jolla, CA 92037. E-mail: bruce@scripps.edu

Over the past 4 years, such sequencing platforms, including the Roche (454) GS FLX sequencer, Illumina genome analyzer, and Applied Biosystems SOLiD sequencer, have become available. They all have in common the capability to process millions of sequence reads in parallel at relatively low cost and are being applied in a number of ways, from resequencing whole genomes, as in The 1000 Genomes Project and The Exome Project (THE 1000 GENOMES PROJECT CONSORTIUM 2010; NG *et al.* 2009), to *de novo* assembly of bacterial and lower eukaryotic genomes, to analyses of gene expression and epigenetic modifications (reviewed by MARDIS 2007 and METZKER 2010).

Here we have applied whole genome sequencing using the Applied Biosystems SOLiD sequencer coupled with low resolution meiotic mapping using bulk segregation analysis (BSA) to identify the mutation responsible for the *Aoba* phenotype, which is characterized by nephritic syndrome and progressive renal failure. The causative mutation destroys the donor splice site in intron 8 of the *Col4a4* gene on chromosome 1. The *Col4a4* gene has not been targeted in mice, and mutations of the orthologous locus in humans cause Alport syndrome (MOCHIZUKI *et al.* 1994), an inherited disorder of basement membranes containing type IV collagen chains that can cause ocular lesions, sensorineural hearing loss, and end-stage renal failure (KASHTAN

1999). In addition to the mutation in *Col4a4*, nine other homozygous mutations that change coding sense were identified elsewhere in the *Aoba* genome.

MATERIALS AND METHODS

Animals: Mice were housed and bred in The Scripps Research Institute vivarium (La Jolla, CA). C57BL/6J and C57BL/10J mice were purchased from The Jackson Laboratory (Bar Harbor, ME). ENU mutagenesis was performed on C57BL/6J mice as described previously (HOEBE *et al.* 2003). Once the mutation in *Col4a4* was identified, *Aoba* mice were genotyped by sequencing PCR products amplified from genomic DNA using the following primers for PCR: forward, 5'AGCCTTGGGTCAGTCCATAGC and reverse, 5'CCATCC CAGTGAAATGGCTAC; for sequencing: forward, 5'TCCCA AGGAGCCTTGAAATG and reverse, 5'AAATGGCTACAGT GCTTTTA. All studies were performed in accordance with the guidelines established by the Institutional Animal Care and Use Committee of The Scripps Research Institute.

Blood and urine analyses: Blood was collected into serum separator tubes and stored at -70° until blood urea nitrogen levels were measured at the University of California-San Diego Murine Hematology and Coagulation Core Laboratory (La Jolla, CA). Urine was collected into Eppendorf tubes, and 2 μ l was immediately pipetted onto each pad of a chemstrip 5 OB reagent strip for urinalysis (Roche). Results were read after 1 min.

Hearing tests: Auditory brainstem responses were measured as previously described (SCHWANDER *et al.* 2007).

Histology: Whole kidneys were placed in plastic tissue cassettes and fixed for 24 hr in 10% buffered formalin. The tissue samples were moved to 70% ethanol before paraffin embedding and sectioning. Adjacent kidney sections, 3 μ m thick, were processed for hematoxylin and eosin, Periodic acid-Schiff, and Trichrome staining at the University of California, San Diego Histology and Immunohistochemistry Shared Resource.

Whole genome DNA sequencing: A tail sample from a moribund male *Aoba* mouse was digested overnight at 55° in 0.2% SDS lysis buffer and proteinase K. The digested DNA was sheared by passing it through a 20.5-g needle and then subjected to phenol extraction and two room temperature chloroform extractions. After addition of sodium acetate, the DNA was precipitated with 100% ethanol, washed twice with 70% ethanol, and resuspended in nuclease-free water for preparation of the SOLiD library. SOLiD sequencing was performed according to the Applied Biosystems SOLiD 3 System Library Preparation Guide, Templated Beads Preparation Guide, and Instrument Operation Guide.

SOLiD data were analyzed using a Linux computer cluster, with a total of 3936 CPUs. The SOLiD color space raw data were processed using the Applied Biosystems software, Corona Lite. Briefly, the base-space reference sequences were converted to color-space sequences; the 50 color-space raw data were then aligned to the mouse reference genome (NCBI reference assembly build 37) allowing a maximum of 6 mismatches per read. The uniquely matched reads were collected with mismatched locations attached. SNPs and point mutations were identified by scanning a consensus of valid, adjacent, two-color mismatches in the uniquely mapped reads. Total coverage was assessed by counting those coding and/or splicing base pairs covered three times or more by SOLiD. At least three discrepancies from the reference sequence were required at any particular site to identify that site as a potential mutation. Calls were filtered as described in RESULTS, and only

those calls that were located in coding region and were predicted to cause amino acid changes or that were located in splice junctions (≤ 10 nucleotide distance from the exon boundary) were collected for interrogation by capillary sequencing.

Estimation of the sensitivity of heterozygote detection: A discrepancy is called by Corona Lite if three or more reads are not in agreement at a point corresponding to a particular base within the reference sequence. A binomial formula may be used to estimate the likelihood of heterozygote detection as a function of depth of coverage. For example, if a base is heterozygous and has been covered with three reads, the chance of a discrepant call is 0.125. If the base has been covered four times, the chance of a discrepant call is 0.3125. For all bases corresponding to coding sequences or splice junctions that were covered three or more times, the binomial probability was used to estimate the likelihood of heterozygote detection. By summing the likelihood of detection at all individual nucleotides, a composite likelihood of detection was determined for the entire target region.

Validation sequencing: For validation sequencing, genomic DNA from the index mouse analyzed by SOLiD sequencing was amplified by PCR with primers that were designed using a Perl script embedded with the Prime program from the GCG DNA software analysis package. The PCR products were purified on a Biomek FX using AMPure beads (Agencourt) and sequenced using Big Dye Terminator on an ABI 3730 XL capillary sequencer. Evaluation of validation sequencing data was performed using a Perl script embedded with Phred-Phrap, and discrepant base pair calls were visualized with Consed.

Reverse-transcriptase PCR: Kidneys from age- and sex-matched C57BL/6J and *Aoba* mice homozygous for the *Col4a4* mutation were collected in RNAlater solution (Ambion) and stored at -70° . Organs were thawed at room temperature, and total RNA was prepared using TRIzol (Invitrogen). The RNA was reverse transcribed using the RETROscript kit (Ambion), and cDNA was amplified using JumpStart REDTaq ReadyMix reaction mix (Sigma-Aldrich) with two different sets of PCR primers. Set 1 forward (5'CTGGATTTCAGGTGTGGATG) and reverse (5'AACCCAGGGCTTCCAGGATGA) primers yield a 346-bp product. Set 2 forward (5'TGCTCTGTCTGCCA GTGTTTT) and reverse (5'CCCCTTTGTCCCTTCAGTCCA) primers yield a 575-bp product. The PCR products were purified using a QIAquick PCR purification kit (Qiagen) and sequenced on an ABI 3730 XL DNA analyzer using two different sets of primers: set 1, forward 5'ACAATGGCTCAC GTGGTGAT and reverse 5'AGGATGACCTATAGGTCCTGC; set 2, forward 5'CTGCCAGTGTTTTTCCTGAGAA and reverse 5'AGTGGCAGGATTTCCCTTCAA. Trace alignments were made using PhredPhrap and visualized with Consed.

Bulk segregation analysis: Homozygous *Aoba* males were outcrossed to C57BL/10J females, and the F₁ progeny were intercrossed to generate F₂ mice for mapping. Proteinuria was measured in 4- to 5-month-old F₂ mice. BSA was performed as described (XIA *et al.* 2010) for a total of 17 F₂ mice with proteinuria and 28 F₂ mice without protein or trace amounts of protein in their urine.

RESULTS

Isolation of the *Aoba* pedigree: We initially isolated the *Aoba* pedigree in a forward genetic screen for mutations that impair the innate immune response to sublethal murine cytomegalovirus (MCMV) infection. A single G3 mouse succumbed on day 6 after infection with

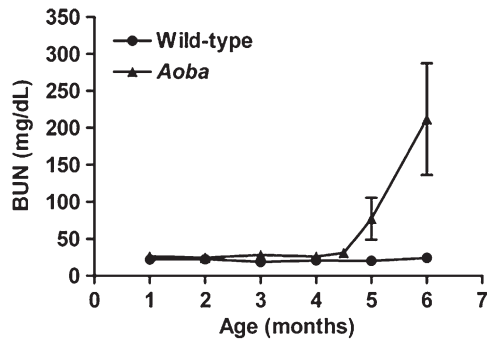


FIGURE 1.—Elevated blood urea nitrogen levels in *Aoba* mice. Levels of blood urea nitrogen in wild-type and *Aoba* mice at the indicated ages. Each point represents the mean \pm SEM for 3–10 mice per group, except for the data for 1-month-old mice, which was determined for only one mouse of each genotype.

10^5 PFU of MCMV (Smith strain), an inoculum that does not cause death or discernible illness in wild-type C57BL/6J mice (CROZAT *et al.* 2006). The G3 siblings of the index mouse were randomly intercrossed in an attempt to fix the mutation that we presumed compromised innate immunity. In the process of breeding *Aoba* mice under SPF conditions, we observed that an unusually high frequency of mice within this pedigree became moribund and eventually died between 6 and 7 months of age. A homozygous stock of mice with early mortality was established after five generations of random sibling crosses.

Progressive renal disease and focal segmental glomerulosclerosis in *Aoba* mice: Gross examination of moribund *Aoba* mice revealed that they had small, pale kidneys and enlarged stomachs filled with undigested food; the kidneys of younger siblings appeared normal. In addition, the blood urea nitrogen (BUN) levels were elevated in 5-month-old and moribund *Aoba* mice, but not in younger animals (Figure 1). Proteinuria (Table 1), hematuria (Table 2), and leukocytes in the urine (Table 3) could be detected in *Aoba* mice as early as 2 months of age. These observations suggested that *Aoba* mice develop early nephritic syndrome and progress to end-stage renal failure between 6 and 7 months of age.

To determine the cause of renal failure in *Aoba* mice, the kidneys of 2- and 5-month-old animals were analyzed by histology. In a 5-month-old *Aoba* mouse, there were regions of subcapsular scarring, evidenced by indentations and pitting of the cortical regions of the kidney with apparent concentration or crowding of glomeruli caused by tubular atrophy and interstitial fibrosis with dilated tubules in some areas (Figure 2A). Trichrome staining highlighted areas of collagen deposition, indicative of scarring, which extended from the cortex into the medulla in many areas of the kidney (Figure 2, D and E). PAS staining showed focal segmental glomerulosclerosis and in some areas, the sclerosis of whole glomeruli (Figure 2B). Glomeruli in the apparently unaffected areas of kidney from a 2-month-old *Aoba*

TABLE 1
Proteinuria in wild-type and *Aoba* mice

Protein ^a	2 mo	3 mo	4 mo	5 mo ^b
Negative	1	14	0	0
Trace	3	6	0	1
(+) or 30 mg/dl	1	2	5	1
(++) or 100 mg/dl	0	1	2	2
(+++ or 500 mg/dl	0	0	0	0
Total no. of mice analyzed	5	23	7	4

^a All wild-type mice (2–5 per age group, 14 total) were negative for proteinuria.

^b Urine could not be collected from 2 additional 5-month-old mice.

mouse showed occasional evidence of increased matrix with the PAS stain but most were normal (supporting information, Figure S1A). Trichrome staining of the kidney from the younger *Aoba* mouse did not reveal major differences from wild-type sections with the exception of some dilated tubules and a small depression on the surface of the kidney with collagen deposition (Figure S1C).

Identification of the *Aoba* mutation by whole genome sequencing: To identify the mutation responsible for renal disease in this pedigree, we used the Applied Biosystems SOLiD 3 sequencing platform to sequence unfractionated genomic DNA obtained from a moribund *Aoba* male which had been inbred for six generations. Briefly, genomic DNA was sheared into 100- to 110-bp fragments, ligated to P1 and P2 adaptors, and amplified using library primers. The resulting short fragment DNA library was then clonally amplified onto SOLiD P1 DNA beads. Following an enrichment step to remove non-templated beads, 640,598,664 templated beads were loaded onto two slides and subjected to SOLiD sequencing through an extension length of 50-bp fragment reads. A total of 639,041,075 beads yielded usable data; 311,149,104 reads (48.69% of the usable beads) were mapped to the reference genome (C57BL/6J assembly build 37) with a maximum of six mismatches allowed for each read. A total of 222,181,879 reads (34.77% of the usable beads, bearing 11.1 Gbp of DNA sequence) were uniquely mapped to the reference genome (Table 4),

TABLE 2
Hematuria in wild-type and *Aoba* mice

Blood ^a	2 mo	3 mo	4 mo	5 mo
Negative	1	10	2	0
Trace	4	8	4	2
~50 Ery/ μ l	0	4	0	0
~250 Ery/ μ l	0	0	0	2
Total no. of mice analyzed	5	22	6	4

^a All wild-type mice (2–5 per age group, 13 total) were negative for hematuria.

TABLE 3

Presence of leukocytes in urine from wild-type and *Aoba* mice

Leukocytes ^a	2 mo	3 mo	4 mo	5 mo
Negative	0	8	0	0
Trace	2	8	0	1
(+)	3	4	5	3
(++)	0	2	1	0
Total no. of mice analyzed	5	22	6	4

^aAll wild-type mice (2–5 per age group, 13 total) were negative for the presence of leukocytes in the blood.

covering a total of 2,095,151,211 bp or 78.92% of the 2,654,895,218-bp haploid mouse genome with at least one read. After subtracting mismatches, we determined that 2,088,564,461 bp (78.67% of all base pairs in the reference genome) were covered at least once; 1,848,722,789 bp (69.63%) were covered at least twice; and 1,600,276,098 bp (60.28%) were covered at least three times. Substantially higher coverage was observed across the coding region of the reference genome derived from the “OR” set of NCBI and Ensembl annotations (*i.e.*, among all coding nucleotides or splice junctions annotated by either NCBI or Ensembl). After subtracting mismatches, 37,299,693 bp of the 41,979,949 bp of coding region (88.85%) were covered at least once; 34,593,375 bp (82.40%) were covered at least twice; and 31,301,063 bp (74.56%) were covered at least three times (Table 5).

Among those nucleotides covered three or more times by SOLiD reads, a total of 29,614 discrepancies from the reference sequence were identified across the genome. Three levels of filtration were applied to these discrepancies (Table 6). First, we ignored discrepancies that were also observed in a total of seven SOLiD plate runs executed on DNA samples obtained from mice unrelated to *Aoba*, reasoning that most of these calls represented systematic errors in sequencing or the effect of false alignments to repetitive DNA. This filter eliminated

18,580 discrepancies, leaving 11,034 discrepancies for further consideration. Second, we ignored discrepancies that did not fall within coding sequences or splice junctions (defined as those 10 bp of intronic sequence adjacent to exons). This eliminated 10,846 discrepancies, leaving 188 for further consideration. Finally, we ignored changes within coding sequence that did not alter coding sense: 42 of the remaining 188 calls fell within this category and were eliminated from consideration, leaving 146 putative mutations (64 putative homozygous changes and 82 putative heterozygous changes) to be tested by capillary sequencing. Of these 146 putative mutations, 135 fell within coding region and changed coding sense, and 11 fell within splice junctions and could potentially alter coding sense.

Primers were designed to amplify and sequence all 146 putative mutations from both wild-type and *Aoba* genomic DNA using a Beckman FX robot and an ABI 3730 XL capillary sequencer. Of these, 132 of the nucleotides (90.41%) were successfully interrogated on first pass: 10 putative mutations (7.6%) were validated as true homozygous mutations and 122 were unambiguously excluded as false positives. None of the heterozygous mutations were confirmed by capillary sequencing. The false discovery rate for homozygous mutations was ~ 1 for every 580,000 bases, and for both homozygous and heterozygous mutations, ~ 1 for every 260,000 bases covered three or more times by SOLiD.

Among the 10 validated homozygous mutations, 8 altered coding sense and 2 were located within splice junctions (Table 7). Of the latter, one was a G→A transition in the conserved donor splice site in the eighth intron of *Col4a4* on chromosome 1. This mutation resulted in aberrant transcription of the *Col4a4* gene with deletion of 8 nucleotides, a frameshift error, the addition of 13 aberrant amino acids after residue 176, and a premature stop codon (Figure 3).

Mapping the *Aoba* phenotype using bulk segregation analysis: We hypothesized that the mutation in *Col4a4* was the most likely cause of renal pathology in *Aoba*

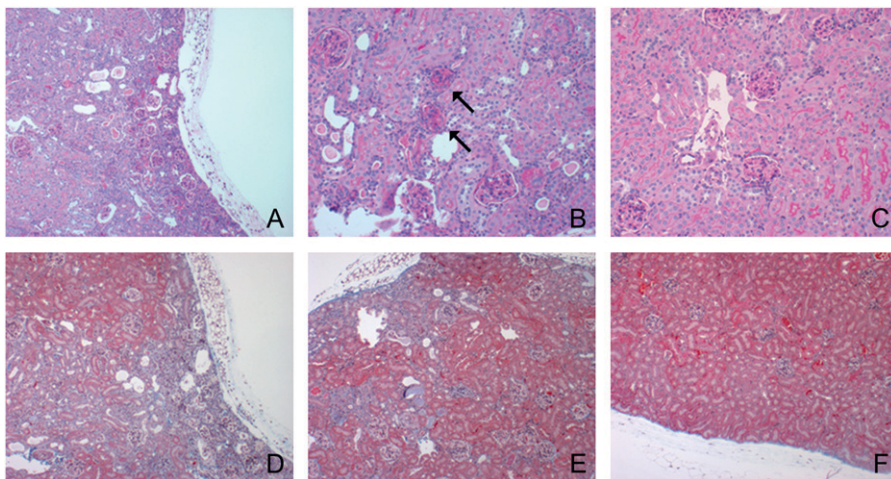


FIGURE 2.—Focal segmental glomerulosclerosis in *Aoba* mice. Serial kidney sections, 3 μ m thick, were stained with PAS (A–C) or Trichrome (D–F) and imaged at $\times 100$ (A and D) or $\times 200$ (B and C, E and F). (A) Pitting of the cortical surface, crowding of glomeruli, tubular atrophy, and dilated tubules in 5-month-old *Aoba* mouse. (B) Completely sclerosed glomeruli (arrows) in 5-month-old *Aoba* mouse. (C) Normal staining in 5-month-old wild-type mouse. (D) Extensive collagen deposition and scarring in 5-month-old *Aoba* mouse. (E) Scarring extending into medulla in 5-month-old *Aoba* mouse. (F) Normal staining in 5-month-old wild-type mouse.

TABLE 4
Summary of SOLiD sequencing statistics

	Templated beads	Usable beads	Total mapped reads	Uniquely mapped reads
<i>Aoba</i>	640,598,664	639,041,075	311,149,104	222,181,879 ^a

^a The value indicated is the total number of reads mapped to the reference genome (C57BL/6J assembly build 37) without subtracting mismatches. The total number of uniquely mapped reads gave 11.1 Gbp of DNA sequence data.

mice. *Col4a4* encodes the type IV collagen $\alpha 4$ chain, which, along with type IV collagen $\alpha 3$ and $\alpha 5$ chains, makes up the basement membranes in the tubules and glomeruli of adult kidneys (KASHTAN and KIM 1992; MINER and SANES 1994). Mutations that prevent the expression of any one of these three type IV collagen chains in humans cause Alport syndrome, characterized by proteinuria, focal segmental glomerulosclerosis, and end-stage renal failure (BARKER *et al.* 1990; LEMMINK *et al.* 1994; MOCHIZUKI *et al.* 1994). Mutations in type IV collagen $\alpha 5$ in dogs (ZHENG *et al.* 1994) and mutations of $\alpha 3$ (COSGROVE *et al.* 1996), $\alpha 3$ and $\alpha 4$ (LU *et al.* 1999), or $\alpha 5$ (RHEAULT *et al.* 2004) in mice also result in glomerular basement membrane abnormalities that culminate in end-stage renal disease.

We excluded *Col4a3* and *Col4a5* as candidates by sequencing their entire coding regions and all other candidate genes by genetic mapping. For the latter purpose, homozygous *Aoba* males were outcrossed to C57BL/10J females, and the F₁ progeny were intercrossed to yield F₂ mice. We measured proteinuria in 4- to 5-month-old F₂ animals: 17 of 45 F₂ mice tested had high levels of protein in their urine, whereas the remaining 28 had no detectable protein or trace amounts. Leukocytes were also detected in the urine of mice with proteinuria, but not in those animals with normal urine samples. Genomic DNA was isolated from the tails of all 17 F₂ mice with proteinuria and 28 normal F₂ mice and quantitated by real-time PCR. Equal amounts of DNA from each mutant and wild-type mouse were added to

separate pools and subjected to BSA, using a total of 127 SNPs that distinguish the C57BL/6J and C57BL/10J strains (XIA *et al.* 2010).

BSA localized the *Aoba* mutation to an ~43 Mb region on chromosome 1, with LOD scores of 7.1, 8.9, and 10 at markers 75,483,331; 95,571,814; and 118,565,405, respectively (Figure 4A). Because the mutation in *Col4a4* is located at position 82,532,315 on chromosome 1, the peak linkage to the marker at 118,565,405 bp was unexpected, but most likely due to the fact that LOD scores calculated from BSA data are based on an estimation of allele frequency within a pool of DNA from multiple mice (XIA *et al.* 2010). To determine exact allele frequencies in the mutant and control groups, we genotyped individual F₂ mice for the C57BL/6J and C57BL/10J alleles of each of these markers (Figure 4B). The LOD scores calculated from this analysis were 5.9, 10.2, and 10.2 at markers 75,483,331; 95,571,814; and 118,565,405, respectively. We conclude, on the basis of these results, that the *Col4a4* mutation on chromosome 1 is responsible for kidney disease in *Aoba* mice.

Hearing in *Aoba* mice: In addition to kidney disease, ~50% of human patients with Alport syndrome have progressive high-frequency sensorineural hearing loss (MYERS and TYLER 1972). Mice with a targeted mutation in *Col4a3* have defects in the cochlear basement membrane accompanied in some animals by minor changes in auditory responses with age (COSGROVE *et al.* 1998). To examine whether *Col4a4* disruption causes sensorineural

TABLE 5
Summary of SOLiD sequencing coverage

Coverage	Total genome in bp (%) ^a	Coding regions and splice junctions in bp (%) ^c
At least one time	2,088,564,461 (78.67) ^b	37,299,693 (88.85)
At least two times	1,848,722,789 (69.93)	34,593,375 (82.40)
At least three times	1,600,276,098 (60.28)	31,301,063 (74.56)

^a Percentage of coverage of the total genome was determined by dividing the number of base pairs covered by the estimated size of the haploid mouse genome: 2,654,895,218 bp.

^b A total of 222,181,879 total reads was uniquely mapped to the reference genome. Coverage of 1× is the number of uniquely mapped reads with mismatches subtracted.

^c Percentage of coverage of the total coding regions and splicing junctions was determined by dividing the number of base pairs covered by 41,979,949 bp, the estimated size of the coding region of the reference genome derived from the "OR" set of NCBI and Ensembl annotations. Mismatches were subtracted from the values shown. The splicing junction sequence is defined as 10 nucleotides after the previous exon or before the next exon.

TABLE 6

DISCUSSION

Overview of filters applied to SOLiD sequencing data set	
3× coverage of total genome ^a	1,600,276,098 bp
Total no. of discrepancies ^b	29,614
Discrepancies unique to <i>Aoba</i> run ^c	11,034
Discrepancies in splice sites or coding regions	188
Discrepancies in splice sites or coding regions that change coding sense	146 (11 in splice sites; 135 in coding region)
Discrepancies successfully reanalyzed by capillary sequencing	132
Confirmed mutations ^d	10 (2 in splice sites; 8 in coding region)

^aTotal number of base pairs of whole genome covered at least 3× with mismatches subtracted.

^bTotal number of differences between SOLiD sequencing data and reference data.

^cDiscrepancies shared between the *Aoba* run and a total of seven other SOLiD sequencing runs performed on unrelated mice were discarded.

^dAll confirmed mutations were homozygous.

hearing loss in mice, we determined auditory thresholds by measuring the auditory brainstem responses (ABR) of 5-month-old *Aoba* and age-matched control animals. Broadband click stimuli were applied to mice starting at 90 dB and then at progressively decreased sound intensities. We determined the average auditory threshold to be 48 ± 4 dB in 5-month-old C57BL/6J mice. Of six 5-month-old *Aoba* mice tested, four had auditory thresholds within the range determined for age-matched control mice; however, two mice had elevated thresholds of 70 dB (Figure 5, A and B). Similar analysis of the F₂ mice used for mapping proteinuria revealed that 3 of the 17 mutant mice had elevated thresholds of >70 dB (data not shown). These results suggest that the *Aoba* mutation in *Col4a4* may result in moderate sensorineural hearing loss with incomplete penetrance.

Here we have used whole genome sequencing based on the Applied Biosystems SOLiD 3 sequencing platform coupled with coarse genetic mapping by BSA to identify an ENU-induced mutation in *Col4a4* responsible for renal failure in mice. The mutation occurred in the conserved donor splice site in the eighth intron of the *Col4a4* gene on chromosome 1. It causes a frame-shift error in the *Col4a4* transcript that most likely abolishes expression of the *Col4a4* gene product, type IV collagen $\alpha 4$. Loss of *Col4a4* expression almost certainly underlies the proteinuria, BUN elevation, and focal segmental glomerulosclerosis that culminate in kidney failure at 6–7 months of age in *Aoba* mice. Notably, *Col4a3*^{-/-} mice (C57BL/6J background) progress to end-stage renal disease in approximately the same period of time (ANDREWS *et al.* 2002).

In mice, the type IV collagen $\alpha 4$ chain is highly expressed in kidney and lung and detected at lower levels in heart, muscle, and skin (MINER and SANES 1994). Within the kidney, type IV collagen $\alpha 4$ forms heterotrimers with the type IV collagen $\alpha 3$ and $\alpha 5$ chains; these trimers assemble into a network that comprises the tubular and glomerular basement membranes (KASHTAN and KIM 1992; MINER and SANES 1994). Mutations that abolish the expression of any one of these collagen chains prevent expression of the other two, resulting in abnormal collagen chain composition of the glomerular basement membranes, such as persistence of an embryonic-like $\alpha 1/\alpha 1/\alpha 2$ collagen network (MINER and SANES 1994). Although the function of the glomerular basement membranes is initially retained, differences in the physical and biological properties of the $\alpha 3/\alpha 4/\alpha 5$ and $\alpha 1/\alpha 1/\alpha 2$ collagen networks likely underlie the progressive loss of glomerular function in humans and animals with Alport syndrome (KALLURI *et al.* 1997; GUNWAR *et al.* 1998). In humans, Alport syndrome presents with considerable clinical heterogeneity, and genetic modifier loci mapped to chromosomes 9 and 16 influence the extent and onset of disease in mice (ANDREWS *et al.* 2002).

TABLE 7

Summary of confirmed homozygous mutations identified in the *Aoba* genome

Chr	Chromosome location	Gene	Location of mutation	Nucleotide change	Type of mutation	Predicted effect
1	82,449,000–82,583,000 bp(-)	<i>Col4a4</i>	82532315	G→A	N/A	Splice donor site
1	16,043,393–16,044,001 bp(+)	<i>Gm7634</i>	16043941	T→A	L144Q	Missense
3	92,304,185–92,304,415 bp(-)	<i>Spr4</i>	92304265	C→T	Q51*	Nonsense
5	76,913,356–76,945,609 bp(+)	<i>Gm7271</i>	76945517	T→C	S143P	Missense
6	145,165,227–145,198,751 bp(-)	<i>Kras</i>	145180734	A→T	M111L	Missense
6	94,554,523–94,650,152 bp(-)	<i>Lrig1</i>	94559081	A→G	T707A	Missense
6	43,123,951–43,125,004 bp(+)	<i>Olfir13</i>	43124558	C→T	T191I	Missense
6	64,992,661–65,066,043 bp(+)	<i>Smarcad1</i>	65024908	T→A	F344I	Missense
14	103,512,630–103,746,017 bp(-)	<i>Mycbp2</i>	103623679	T→C	n/a	Splice acceptor site
18	5,591,872–5,774,908 bp	<i>Zeb1</i>	5767074	A→T	H529L	Missense

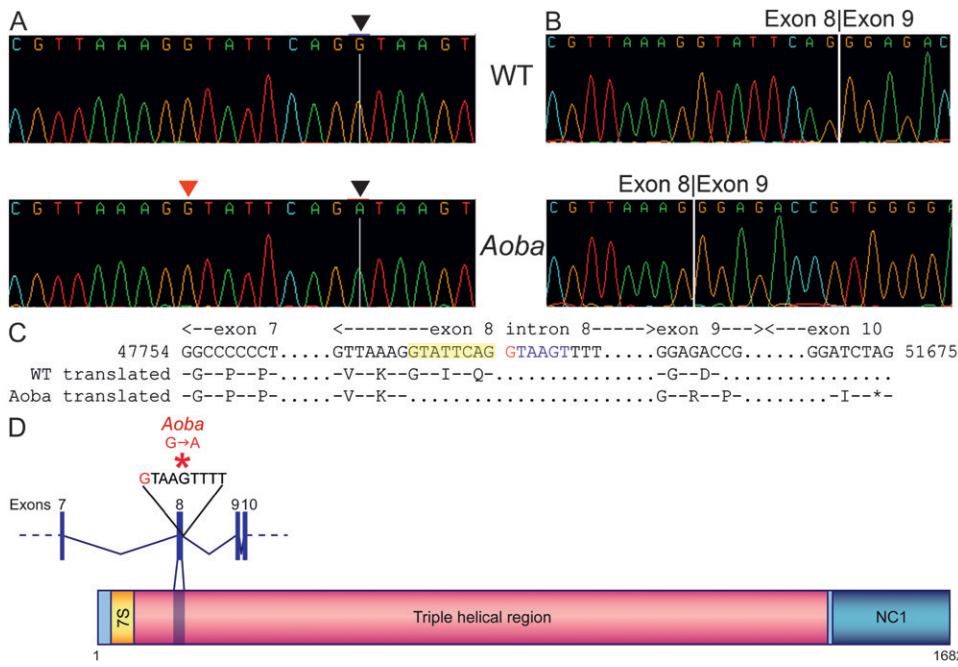


FIGURE 3.—*Aoba* mice have a mutation in *Col4a4*. (A) A single nucleotide transition (G→A, marked with black arrowheads above each trace file) destroys the splice donor site in intron 8 of the *Col4a4* gene. The alternative splice donor site used in the *Aoba Col4a4* gene is marked with a red arrowhead. (B) Reverse transcriptase PCR across the junction between exons 8 and 9 reveals a deletion of eight nucleotides in the *Aoba Col4a4* transcript. The normal and altered junctions between exons 8 and 9 are indicated above each trace file. (C) Schematic highlighting the eight nucleotides deleted from the *Aoba Col4a4* transcript in yellow. The deletion results in a frameshift error and premature stop codon in *Col4a4*, indicated by the translated sequences for wild type and *Aoba Col4a4*. (D) Schematic showing the site of the premature stop codon in the *Col4a4* collagenous domain in *Aoba* mice.

Aoba represents a mouse model in which *Col4a4* has been mutated to the exclusion of other *Col4* family members, although a spontaneous deletion affecting both *Col4a3* and *Col4a4* has been reported elsewhere (LU *et al.* 1999). As such, *Aoba* may serve as a useful animal model for further studies of the pathology of kidney disease and influence of genetic loci on the onset and progression of Alport syndrome. Given our observation that some *Aoba* mice develop sensorineural hearing

loss, this model may help establish whether hearing loss associated with Alport syndrome results from structural defects in the inner ear or ototoxic effects of severe kidney disease (McDONALD *et al.* 1978; COSGROVE *et al.* 1998).

In addition to the mutation in *Col4a4*, nine other homozygous coding changes were identified in the genome of the *Aoba* mouse that we selected for sequencing. Among these, three are predicted by the program Polyphen to be possibly (*Mycbp2* and *Smarcad1*) or probably

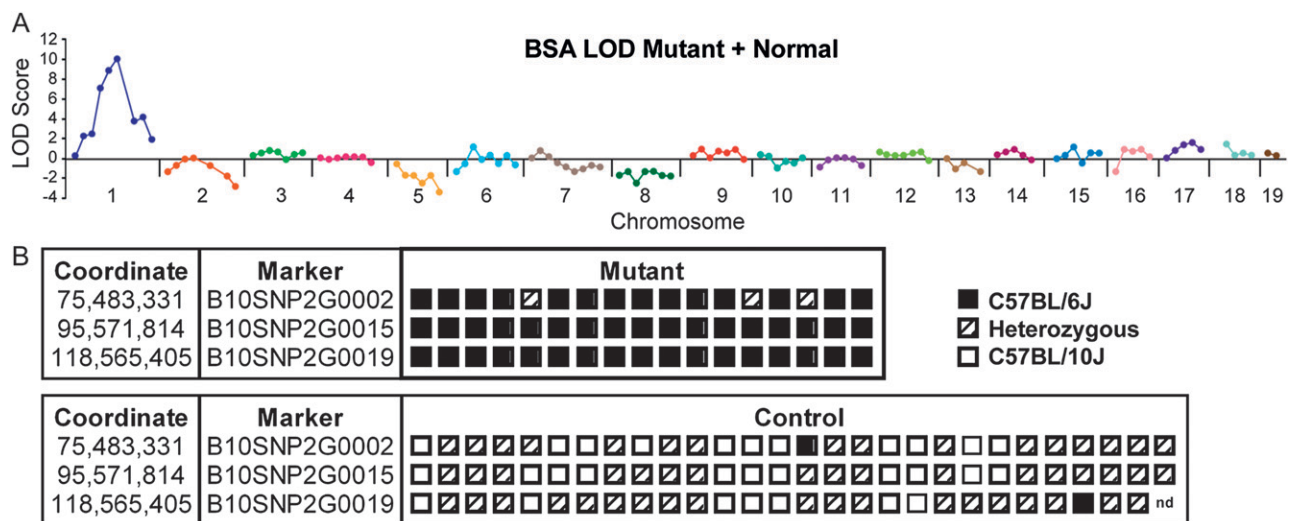


FIGURE 4.—Proteinuria was mapped to chromosome 1 using bulk segregation analysis. (A) Homozygous *Aoba* males were outcrossed to C57BL/10J females, the F₁ progeny were intercrossed, and proteinuria was measured in 4- to 5-month-old F₂ mice. BSA was performed with genomic DNA from 17 mutant and 28 control F₂ mice. The combined LOD score vs. 124 autosomal SNP markers that distinguish C57BL/6J and C57BL/10J mice is shown. The names of the chromosomal markers used for mapping were omitted for clarity, but can be accessed in (XIA *et al.* 2010). (B) Individual mutant and control F₂ mice were genotyped for the C57BL/6J and C57BL/10J alleles of the peak markers on chromosome 1. The C57BL/6J and C57BL/10J alleles are represented in black or white, respectively, and heterozygosity is represented by striped boxes.

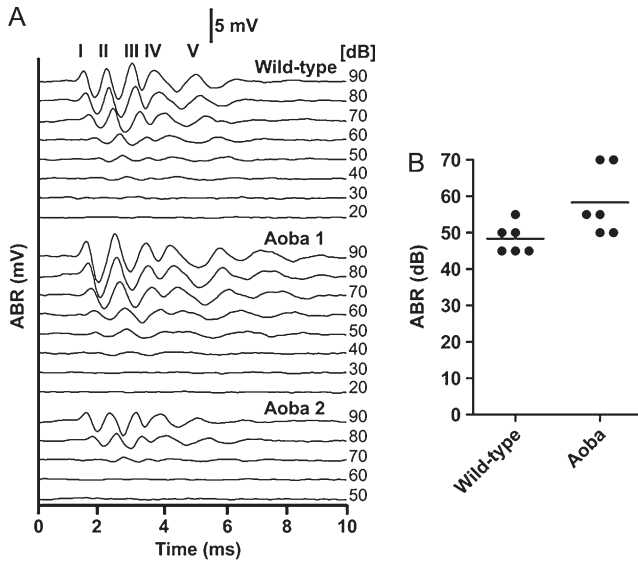


FIGURE 5.—Sensorineural hearing loss in *Aoba* mice. (A) Representative ABR recordings in response to click stimuli in a 5-month-old wild-type mouse and two 5-month-old *Aoba* mice, one of which (mutant 1) had a normal ABR threshold and one of which (mutant 2) had an elevated ABR threshold. (B) Scatter plot showing the ABR thresholds for individual 5-month-old wild-type and *Aoba* mice on a pure C57BL/6J background. Each point represents data from one mouse, and the bar indicates the mean of all values.

(*Zeb1*) deleterious, while one specifies a premature stop codon that likely creates a null allele (*Sprr4*). It is notable that all of the validated coding changes in this animal were homozygous and that none of the 82 putative heterozygous coding changes were confirmed by capillary sequencing. Because *Aoba* mice were subjected to six generations of sibling inbreeding, $\sim 87\%$ of all coding changes that were heterozygous in the founding G2 female of the mutant pedigree would have become fixed either in wild-type or mutant alloforms in the absence of selection favoring the maintenance of heterozygosity (WRIGHT 1921). Given that 10 homozygous coding changes were observed with 74.56% three times or greater coverage by SOLiD with 90.41% validation efficiency, we estimate that ~ 14.8 homozygous coding changes were likely present in the entire coding and splicing target examined.

Of these 14.8 estimated mutations, approximately two-thirds (or 9.9 coding changes) would presumably have been derived from mutations that were transmitted from the G1 sire to the G2 dam and approximately one-third (or 4.9 coding changes) would presumably have been derived from mutations in the G1 sire that were not transmitted to the G2 dam but were transmitted subsequently to the G3 index mouse in the G1 \times G2 cross (Figure S2). The recovery of 9.9 coding changes derived from the G2 dam in homozygous form would suggest that an equal number of mutations once extant in the G2 were lost through fixation of the wild-type allele. Hence,

~ 19.8 heterozygous coding changes once existed in the G2 ancestor of the sequenced mouse and were fixed as wild-type or mutant alleles in the G3 index mouse. Because only 87% of all heterozygous alleles would be expected to achieve fixation in six generations of inbreeding, we estimate that ~ 23 heterozygous coding changes were originally present in the G2 mouse genome and that ~ 3 of these mutations remained unfixed (but undetected) in the G8 mouse that was sequenced. We further infer that about twice as many coding changes—45—were transmitted to the G1 mouse in the sperm derived from the mutagenized G0 male. The occurrence of 45 changes in coding sense within a target comprising ~ 42 Mbp of DNA is consistent with a mutation rate of ~ 1.4 per Mbp as only $\sim 76\%$ of random nucleotide changes lead to a change in coding sense. This is similar to the rate of ENU-induced heritable mutations estimated elsewhere (TAKAHASHI 2007). Thus, coarse mapping coupled with light whole genome sequencing (as performed in the present study) should be sufficient to resolve the causative mutation in the majority of cases.

As estimated above, approximately three unfixed (heterozygous) ENU-induced coding changes may have been present in the genome of the *Aoba* mouse that was sequenced. However, only $\sim 46\%$ of such mutations would be detected within the annotated coding sequence and splice junctions by sequencing to the depth that was performed (see MATERIALS AND METHODS). The chance that none of the three putative unfixed mutations would be detected is therefore $\sim 0.54^3$ or $P = 0.16$.

In summary, whole genome sequencing using the Applied Biosystems SOLiD Analyzer was applied here to identify the mutation responsible for renal disease in *Aoba* mice. Our results suggest a general strategy for finding ENU-induced mutations in mice in which a phenotype is coarsely mapped to a chromosome and the causative mutation is identified by validating discrepancies identified by whole genome sequencing. One caveat to our approach is that BSA is a low-resolution technique, and on occasion, more than one candidate mutation may fall within the critical region: transgenesis or other methods may be necessary to confirm cause and effect. A major advantage, however, in addition to the time and expense saved by obviating the need for fine mapping, is that we are now able to build a repository of mutations that may be rederived by *in vitro* fertilization or intracytoplasmic sperm injection by archiving sperm from all of the male mice whose genomes have been sequenced in the laboratory. Our progress in archiving sperm and annotating mutations is continuously updated at <http://mutagenetix.scripps.edu>, and germline material will be made available to accredited laboratories.

The authors thank Nissi Vardki and the University of California-San Diego Histology and Immunohistochemistry Shared Resource; Mercedes Gutierrez and the Department of Animal Resources for animal husbandry; Jerrica Jaycox and Marlin Jimenez for performing the hearing tests; and Diantha LaVine for artwork in Figures 3 and 4.

This work was funded by Broad Agency Announcement, BAA contract no. HHSN272200700038C (to B.B.) and National Institutes of Health grants DC005969 and DC007704 (to U.M.). C.N.A. is supported by an Irvington Institute Fellowship of the Cancer Research Institute. While this manuscript was being written, a spontaneous *Col4a4* mutation was identified in *bwk* (bilateral wasting kidney) mice (KORSTANJE *et al.* 2009).

LITERATURE CITED

- ANDREWS, K. L., J. L. MUDD, C. LI and J. H. MINER, 2002 Quantitative trait loci influence renal disease progression in a mouse model of Alport syndrome. *Am. J. Pathol.* **160**: 721–730.
- BARKER, D. F., S. L. HOSTIKKA, J. ZHOU, L. T. CHOW, A. R. OLIPHANT *et al.*, 1990 Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* **248**: 1224–1227.
- COSGROVE, D., D. T. MEEHAN, J. A. GRUNKEMEYER, J. M. KORNAK, R. SAYERS *et al.*, 1996 Collagen COL4A3 knockout: a mouse model for autosomal Alport syndrome. *Genes Dev.* **10**: 2981–2992.
- COSGROVE, D., G. SAMUELSON, D. T. MEEHAN, C. MILLER, J. MCGEE *et al.*, 1998 Ultrastructural, physiological, and molecular defects in the inner ear of a gene-knockout mouse model for autosomal Alport syndrome. *Hear Res.* **121**: 84–98.
- CROZAT, K., P. GEORGEL, S. RUTSCHMANN, N. MANN, X. DU *et al.*, 2006 Analysis of the MCMV resistome by ENU mutagenesis. *Mamm. Genome* **17**: 398–406.
- GUNWAR, S., F. BALLESTER, M. E. NOELKEN, Y. SADO, Y. NINOMIYA *et al.*, 1998 Glomerular basement membrane. Identification of a novel disulfide-cross-linked network of alpha3, alpha4, and alpha5 chains of type IV collagen and its implications for the pathogenesis of Alport syndrome. *J. Biol. Chem.* **273**: 8767–8775.
- HOEBE, K., X. DU, P. GEORGEL, E. JANSSEN, K. TABETA *et al.*, 2003 Identification of Lps2 as a key transducer of MyD88-independent TIR signalling. *Nature* **424**: 743–748.
- KALLURI, R., C. F. SHIELD, P. TODD, B. G. HUDSON and E. G. NEILSON, 1997 Isoform switching of type IV collagen is developmentally arrested in X-linked Alport syndrome leading to increased susceptibility of renal basement membranes to endoproteolysis. *J. Clin. Invest.* **99**: 2470–2478.
- KASHTAN, C. E., 1999 Alport syndrome. An inherited disorder of renal, ocular, and cochlear basement membranes. *Medicine (Baltimore)* **78**: 338–360.
- KASHTAN, C. E., and Y. KIM, 1992 Distribution of the alpha 1 and alpha 2 chains of collagen IV and of collagens V and VI in Alport syndrome. *Kidney Int.* **42**: 115–126.
- KORSTANJE, R., S. COOK, R. BRONSON, B. CLEMENS and M. T. DAVISSON, 2009 Identification of a mutation in *COL4A4* in *BWK* mice, a new model for autosomal recessive Alport Syndrome [abstract]. 23rd International Mammalian Genome Conference, Nov 1–4, La Jolla, CA. IMGC; 2009. Poster nr 77.
- LEMMINK, H. H., T. MOCHIZUKI, L. P. VAN DEN HEUVEL, C. H. SCHRODER, A. BARRIENTOS *et al.*, 1994 Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **3**: 1269–1273.
- LU, W., C. L. PHILLIPS, P. D. KILLEN, T. HLAING, W. R. HARRISON *et al.*, 1999 Insertional mutation of the collagen genes Col4a3 and Col4a4 in a mouse model of Alport syndrome. *Genomics* **61**: 113–124.
- MARDIS, E. R., 2007 The impact of next-generation sequencing technology on genetics. *Trends Genet.* **24**: 133–141.
- MCDONALD, T. J., H. ZINCKE, C. F. ANDERSON and N. T. OTT, 1978 Reversal of deafness after renal transplantation in Alport's syndrome. *Laryngoscope* **88**: 38–42.
- METZKER, M. L., 2010 Sequencing technologies: the next generation. *Nat. Rev. Genet.* **11**: 31–46.
- MINER, J. H., and J. R. SANES, 1994 Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: sequence, distribution, association with laminins, and developmental switches. *J. Cell Biol.* **127**: 879–891.
- MOCHIZUKI, T., H. H. LEMMINK, M. MARIYAMA, C. ANTIGNAC, M. C. GUBLER *et al.*, 1994 Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome. *Nat. Genet.* **8**: 77–81.
- MYERS, G. J., and H. R. TYLER, 1972 The etiology of deafness in Alport's syndrome. *Arch. Otolaryngol.* **96**: 333–340.
- NG, S. B., E. H. TURNER, P. D. ROBERTSON, S. D. FLYGARE, A. W. BIGHAM *et al.*, 2009 Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**: 272–276.
- RHEAULT, M. N., S. M. KREN, B. K. THIELEN, H. A. MESA, J. T. CROSSON *et al.*, 2004 Mouse model of X-linked Alport syndrome. *J. Am. Soc. Nephrol.* **15**: 1466–1474.
- SCHWANDER, M., A. SZZANIECKA, N. GRILLET, J. S. BAILEY, M. AVENARIUS *et al.*, 2007 A forward genetics screen in mice identifies recessive deafness traits and reveals that pejvakin is essential for outer hair cell function. *J. Neurosci.* **27**: 2163–2175.
- TAKAHASHI, K. R., Y. SAKURABA and Y. GONDO, 2007 Mutational pattern and frequency of induced nucleotide changes in mouse ENU mutagenesis. *BMC Mol. Biol.* **8**: 52–61.
- THE 1000 GENOMES PROJECT CONSORTIUM, 2010 A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.
- WRIGHT, S., 1921 Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6**: 124–143.
- XIA, Y., S. WON, X. DU, P. LIN, C. ROSS *et al.*, 2010 Bulk segregation mapping of mutations in closely related strains of mice. *Genetics* **186**: 1139–1146.
- ZHENG, K., P. S. THORNER, P. MARRANO, R. BAUMAL and R. R. MCINNES, 1994 Canine X chromosome-linked hereditary nephritis: a genetic model for human X-linked hereditary nephritis resulting from a single base mutation in the gene encoding the alpha 5 chain of collagen type IV. *Proc. Natl. Acad. Sci. USA* **91**: 3989–3993.

Communicating editor: J. C. SCHIMENTI

GENETICS

Supporting Information

<http://www.genetics.org/cgi/content/full/genetics.110.124586/DC1>

Rapid Identification of a Disease Allele in Mouse Through Whole Genome Sequencing and Bulk Segregation Analysis

**Carrie N. Arnold, Yu Xia, Pei Lin, Charles Ross, Martin Schwander, Nora G. Smart,
Ulrich Müller and Bruce Beutler**

Copyright © 2011 by the Genetics Society of America
DOI: 10.1534/genetics.110.124586

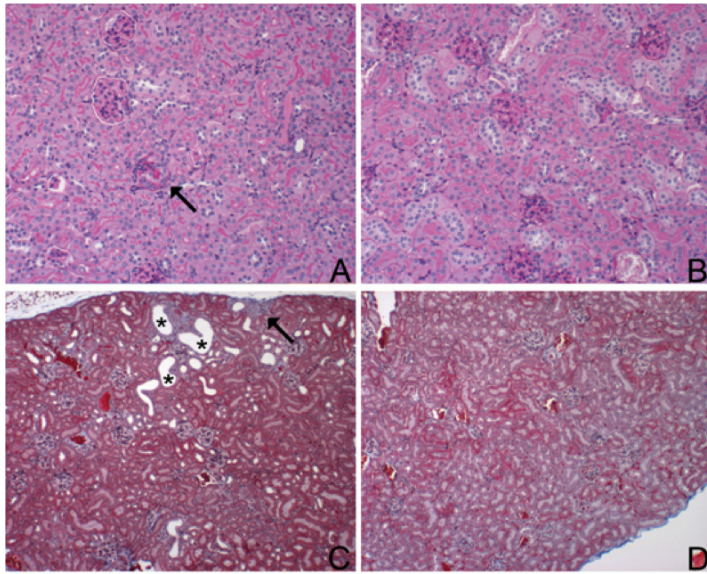


FIGURE S1.—Dilated tubules and scarring in the kidney of a two-month-old *Aoba* mouse. Serial kidney sections, 3 mm thick, were stained with PAS (A-B) or Trichrome (C-D) and imaged at 200X. (A) Increased PAS staining (arrow) in glomerulus in 2-month-old *Aoba* mouse. (B) Normal staining in 2-month old wildtype mouse. (C) Dilated tubules (asterisks) and a region of collagen deposition and scarring (arrow) in 2-month-old *Aoba* mouse. (D) Normal staining in 2-month-old wildtype mouse.

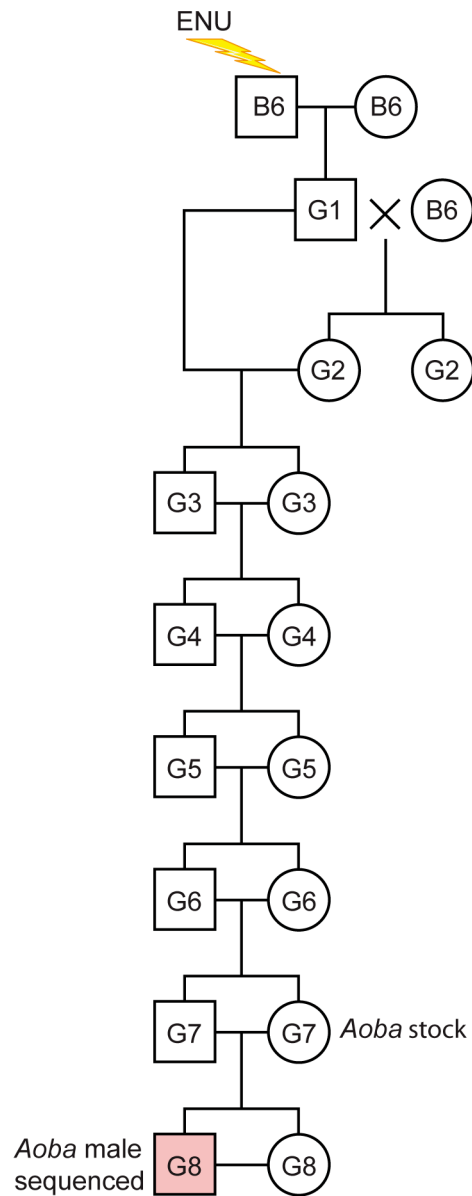


FIGURE S2.—Schematic of the *Aoba* pedigree. An ENU-treated C57BL/6J male was bred to a C57BL/6J female to generate G1 males, which were outcrossed to C57BL/6J females to yield G2 mice. A female G2 mouse was backcrossed to its G1 sire to generate G3 animals, among which, was the original *Aoba* index mouse. Genomic DNA was sequenced from a G8 mouse that developed renal disease (indicated with red shading).