

## Research Article

# Chromosome 1 Sequence Analysis of C57BL/6J-Chr1<sup>KM</sup> Mouse Strain

# Fuyi Xu,<sup>1</sup> Tianzhu Chao,<sup>1</sup> Yiyin Zhang,<sup>1</sup> Shixian Hu,<sup>1</sup> Yuxun Zhou,<sup>1</sup> Hongyan Xu,<sup>2</sup> Junhua Xiao,<sup>1</sup> and Kai Li<sup>1</sup>

<sup>1</sup>College of Chemistry, Chemical Engineering, and Biotechnology, Donghua University, Shanghai, China <sup>2</sup>Department of Biostatistics and Epidemiology, Medical College of Georgia, Augusta University, Augusta, GA, USA

Correspondence should be addressed to Junhua Xiao; xiaojunhua@dhu.edu.cn and Kai Li; likai@dhu.edu.cn

Received 15 December 2016; Revised 9 February 2017; Accepted 15 February 2017; Published 9 April 2017

Academic Editor: Leng Han

Copyright © 2017 Fuyi Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Chinese Kunming (KM) mouse is a widely used outbred mouse stock in China. However, its genetic structure remains unclear. In this study, we sequenced the genome of the C57BL/6J-Chr1<sup>KM</sup> (B6-Chr1<sup>KM</sup>) strain, the chromosome 1 (Chr 1) of which was derived from one KM mouse. With  $36.6 \times$  average coverage of the entire genome, 0.48 million single nucleotide polymorphisms (SNPs) and 96,679 indels were detected on Chr 1 through comparison with reference strain C57BL/6J. Moreover, 46,590 of them were classified as novel mutations. Further functional annotation identified 155 genes harboring potentially functional variants, among which 27 genes have been associated with human diseases. We then performed sequence similarity and Bayesian concordance analysis using the SNPs identified on Chr 1 and their counterparts in three subspecies, *Mus musculus domesticus*, *M. m. musculus*, and *M. m. castaneus*. Both analyses suggested that the Chr 1 sequence of B6-Chr1<sup>KM</sup> was predominantly derived from *M. m. domesticus* while 9.7% of the sequence was found to be from *M. m. musculus*. In conclusion, our analysis provided a detailed description of the genetic variations on Chr 1 of B6-Chr1<sup>KM</sup> and a new perspective on the subspecies origin of KM mouse which can be used to guide further genetic studies with this mouse strain.

#### 1. Introduction

The Chinese Kunming (KM) mouse colony, the largest outbred mouse stock maintained by commercial dealers nationwide in China, has been widely used in pharmaceutical and genetic studies [1]. Unlike other outbred mice, KM mouse has a complex evolutionary history. In 1944 during the World War II, Swiss mice were initially introduced into Kunming, Yunnan Province, China, from the Indian Haffkine Institute by Professor Feifan Tang via the Hump route with the help of the American Volunteer Group [2]. These mice were named KM mice after their initial location in China. Because most other mouse strains were lost and mouse facilities were damaged during the World War II, KM mouse became the only laboratory mouse available afterwards. They were gradually distributed throughout most of the country for medical studies. However, despite the importance of this outbred mouse, its underlying genetic structure remains unclear.

According to the Mouse Genome Informatics (http:// www.informatics.jax.org/), over one thousand quantitative trait loci (QTLs) have been mapped on mouse chromosome 1 (hereafter referred to as Chr 1) including large amounts of QTLs related to metabolism disorder. However, very few candidate genes have been identified partly because of the large QTL intervals. In order to fine map the metabolism disorder QTLs on Chr 1 and identify the candidate genes, we established a population of Chr 1 substitution mouse strains, in which C57BL6/J (B6) was the host strain, and one KM mouse, five inbred strains, and twenty-four wild mice captured from various locations in China were selected as the Chr 1 donors [3]. In order to dissect the genetic structure and variations of this population and better severe further genetic studies, we have resequenced 18 strains of this population including C57BL/6J-Chr1<sup>KM</sup> (B6-Chr1<sup>KM</sup>) with next-generation sequencing technology [4].

In this study, we analyzed the genome sequence data from B6-Chr1<sup>KM</sup> strain and identified 0.48 million single nucleotide polymorphisms (SNPs) and 96,679 indels on Chr 1, of which 6.4% SNPs and 16.3% indels were considered to be novel. Functional annotation suggested that 474 variants had deleterious effect on gene functions. In addition, we explored the KM mouse genetic structure by performing sequence similarity and Bayesian concordance analysis (BCA) on Chr 1. Results suggested that KM mouse was predominately originated from *Mus musculus domesticus* and part of the sequence was from *M. m. musculus*.

#### 2. Materials and Methods

2.1. Animals. B6 and KM mice were purchased from Shanghai SLAC Laboratory Animal Co., Ltd., China. One male KM mouse was mated with female B6 to produce hybrid F1, followed by 8 generations of backcrossing with B6 using marker-assisted selection, then brother×sister mating to create a B6-Chr1<sup>KM</sup> Chr 1 substitution strain [3]. All mice were maintained under specific pathogen-free (SPF) conditions according to the People's Republic of China Laboratory Animal Regulations, and the study was conducted in accordance with the recommendations of and was approved by the Laboratory Animal Committee of Donghua University.

2.2. DNA Sequencing. B6-Chr1<sup>KM</sup> genomic DNA was extracted from tail tissue of a male mouse using an AxyPrep<sup>™</sup> Multisource Genomic DNA Miniprep Kit (Axygen, Hangzhou, China) according to the manufacturer's protocol.

Purified genomic DNA was sheared and size selected (300-500 bp). Paired-end sequencing  $(2 \times 125 \text{ bp})$  was carried out with an Illumina HiSeq 2500 instrument (Illumina Inc., San Diego, CA, USA) on two lanes by WuXi AppTec (Shanghai, China) according to the manufacturer's protocol.

2.3. Read Alignment. Raw reads were filtered using NGS QC toolkit v2.3 [5] to remove reads containing more than 30% low-quality (Q20) bases. Filtered reads were aligned to the C57BL/6J reference genome (December 2011 release of the mouse reference genome (mm10) from Ensembl) using BWA (version 0.7.10-r789) with 12 threads [6]. The resulting SAM file was converted to a binary format and sorted with SAMtools v1.1 [7], followed by the marking of duplicate reads using picard-tools v1.119 (http://picard.sourceforge.net). To improve SNP and indel calling, indel realignment was conducted with Genome Analysis Toolkit (GATK v3.3) [8].

2.4. SNP/Indel Identification and Annotation. SNPs and indels were called using SAMtools mpileup and BCFtools call functions [7], with the '-uf and '-cv' parameters, respectively. To identify a high-quality variant data set, variants were filtered using the BCFtools filter and VCFtools varFilter function [9]. The following parameters were used: for BCFtools filter, '-g 10 -G 3 -i 'QUAL>10 && MIN(MQ)>25 && MIN(DP)>6 && MAX(DP)<199 && (DP4[2]+DP4[3]) > 2', and for VCFtools varFilter, '-2 0'.

Ensembl Variant Effect Predictor tool (VEP, v78) [10] was used to characterize the SNPs and indels, and the algorithm SIFT was used to predict whether a missense variant would have a deleterious effect on a protein-coding gene.

2.5. Sequence Similarity Analysis. SNP information for WSB/EiJ (WSB), PWK/PhJ (PWK), and CAST/EiJ (CAST) was downloaded from the Mouse Genome Project (MGP) database of the Sanger Institute. The Chr 1 consensus sequence for each strain was constructed using the SAMtools consensus parameters. The repeat-masked B6-Chr1<sup>KM</sup> Chr 1 sequence was divided into 1955 100 kb segments. The similarities of each segment with the corresponding segments in the WSB, CAST, and PWK were evaluated. Sliding window similarity analysis was also performed using 500 kb windows and 100 kb sliding intervals.

2.6. *Phylogenetic Analysis*. Phylogenetic analysis was conducted with the previously reported BCA method [11], with the *Rattus norvegicus* Chr 1 sequence (version rn5) downloaded from Ensembl used as the out-group. Briefly, consensus sequences from the WSB, PWK, and CAST strains were mapped to the alignment and gaps filled with Ns. Collinear segments were partitioned into 830 loci using a minimum description length algorithm with a default maximum cost.

2.7. Phylogenetic Tree Evaluation. Nexus files corresponding to the WSB-derived or PWK-derived regions were converted to FASTA files, and then a neighbor-joining phylogenetic tree was constructed using MEGA6 program [12]. Subsequently, 1000 bootstrap replicates were performed to generate branch support values.

#### 3. Results

3.1. B6-Chr1<sup>KM</sup> Genome Background. Chromosome substitution strains, also named as consomic strains, are designed to simplify the genome background and increase the power and speed of QTL mapping. The characteristic of consomic strain is that it only contains a single chromosome from the donor strain substituting the corresponding chromosome in the host strain. For B6-Chr1<sup>KM</sup> consomic strain, Chr 1 sequence was derived from one KM mouse, while the genome background was from the B6 strain (Figure 1). In addition, sequences in the primary mouse reference assembly come from the same B6 strain. Therefore, our analysis of B6-Chr1<sup>KM</sup> whole genome resequencing data only focused on Chr 1.

3.2. SNP and Indel Discovery. In this study, approximately one billion reads from the B6-Chr1<sup>KM</sup> mouse strain were generated on two lanes of Illumina HiSeq 2500. A total of 78.65% of the reads were considered to be clean reads after quality control evaluation. Of them, more than 99% were aligned to the B6 mouse reference genome (mm10) using BWA with a mean genome-wide coverage of 36.6×.

A total of 479,956 SNPs and 96,679 indels were detected using SAMtools/BCFtools on Chr 1, in which 462,755 (96.42%) of the sites were homozygous. These variants were compared with variant calls from 36 key mouse strains from



FIGURE 1: The characteristics of B6-Chr1<sup>KM</sup> genome background. Blue bars represent B6 chromosome while the red represents KM mouse chromosome.



FIGURE 2: Distribution of SNP density on B6-Chr1<sup>KM</sup> Chr 1. The SNP density is represented by the number of SNPs mapped within 100 kb physical intervals across Chr 1.

the Sanger Institute [13] as well as NCBI dbSNP142 variant data sets. This led to the identification of 449,089 SNPs (93.6%) as known, and the remaining 30,867 SNPs (6.4%) were classified as novel. For indels, 15,723 (16.3%) were classified as novel. In addition, we evaluated the variant calls using Sanger sequencing in our previous study which achieved high accuracy with 0.57% false positive and 0% false negative rate [4].

Next, we detected the distribution and density of SNPs over 100 kb window sizes. The observed average SNP density across the entire Chr 1 was 250 per 100 kb. However, different regions showed varying densities. For example, 29.5% of the Chr 1 sequence had an extremely low (0–5 SNPs per 100 kb) SNP density, while 9.1% had a high density (800 or more SNPs per 100 kb). The proximal region of Chr 1 was the longest region with a low SNP density encompassing nearly 25 Mb (Figure 2).

3.3. Functional Consequences of the SNPs and Indels. The putative consequences of SNPs and indels were cataloged using VEP from Ensembl (Table 1). The majority of the SNPs were located in intergenic (224,557, 18.7%) and intronic regions (575,013, 47.8%), and nearly 12% were classified as noncoding transcript variants. With regard to splice sites, 40 splice variants (including splice donor and splice acceptor variants) were found. The numbers of SNPs causing a premature stop codon or stop loss were 19 and 5, respectively. In addition, 2,378 (0.2%) missense variants were detected in

358 genes (one or more variants per gene). Among them, 380 variants (31.6%) from 113 genes were considered to have deleterious effects (SIFT < 0.05). Similar to the SNPs, the majority of indels were intronic (49.3%) and intergenic (17.1%) or within 5 kb upstream or downstream of a gene (16.9%). Only a small number of indels caused frameshift (22) and stop gain or loss (2). Among the novel variants, 7 caused a disruption of the translational reading frame; 10 were predicted as premature truncation of the protein due to gain or loss of stop codons; and 9 were located in splice donor regions. In addition, 104 novel missense variants from 20 genes had deleterious effects.

Next, we annotated these genes containing amino acid altering variants (SIFT < 0.05) and those with stop gain or loss, frameshift, and splice region variant genes with the Human-Mouse: Disease Connection database from Mouse Genome Informatics [14]. This analysis, which contained 155 genes, resulted in 27 genes associated with 49 different human disease-related phenotypes (Table 2), including macular degeneration, breast cancer, and immunodeficiency. Among these 27 disease genes, 9 have been investigated with mouse models, which had an in-depth phenotype information in different mouse genome background.

3.4. Sequence Similarity Analysis. The house mouse, Mus musculus, consists of three principal subspecies, with M. m. domesticus in Western Europe and the Middle East, M. m.

TABLE 1: Predictions of functional consequences of SNPs and indels.

Consequences	SNPs	Novel SNPs	Indels	Novel indels
splice_donor_variant	29	9	2	0
splice_acceptor_variant	11	0	4	0
stop_gained	19	8	1	1
frameshift_variant	0	0	22	7
stop_lost	5	1	1	0
start_lost	11	3	2	0
missense_variant	2378	486	_	0
inframe_insertion	0	0	28	2
inframe_deletion	0	0	26	2
splice_region_variant	1117	63	244	18
synonymous_variant	4238	281	0	0
stop_retained_variant	3	0	0	0
coding_sequence_variant	1	0	1	0
mature_miRNA_variant	4	2	2	0
5_prime_UTR_variant	1210	86	198	31
3_prime_UTR_variant	6563	484	1617	191
non_coding_transcript_exon_variant	11,955	640	2140	290
intron_variant	575,013	36,838	139,815	21,458
NMD_transcript_variant	42,052	2609	10,291	1372
non_coding_transcript_variant	143,110	8770	32,985	5190
upstream_gene_variant	96,888	8357	24,321	3992
downstream_gene_variant	93,184	5752	23,557	3390
intergenic_variant	224,557	13,198	48,568	8312

Consequences were predicted using Ensembl VEP and gene models from Ensembl version 76. Novel SNPs or indels are defined as variants that were not in MGP and dbSNP142 data sets.

musculus in Eastern Europe and Asia, and M. m. castaneus in Southeast Asia and India. Three genome sequences of the wild-derived inbred mouse strains, WSB, PWK, and CAST, which are broadly used to represent each of the subspecies, were selected for phylogenetic analysis. A Chr 1 consensus sequence was constructed for each strain using the SNP information from MGP. Because the simplest way to analyze phylogenetic divergence is by assessing sequence similarity, the Chr 1 sequence was separated into 1955 100 kb blocks and the similarities between each fragment and the corresponding sequences from WSB, PWK, and CAST were determined. The Chr 1 sequence was found to contain a large number of fragments with high sequence similarity to the corresponding sequence in WSB (Figure 3(a)), which is consistent with previous reports showing that KM mouse is derived from Swiss mice originated from the M. m. domesticus subspecies [1]. In addition, a bimodal distribution of blocks with two peaks of similarity was observed in a comparison of B6-Chr1<sup>KM</sup> Chr 1 with PWK counterpart (Figure 3(a)). The first peak had only 99.05–99.1% sequence similarity to PWK, indicating the intersubspecies genome divergence of the Chr 1 sequence from M. m. musculus. The second peak had >99.7% sequence similarity to PWK (Figure 3(a)), indicating that the sequence of *M. m. musculus* introgressed into the KM mouse Chr 1. For the comparison of B6-Chr1<sup>KM</sup> and CAST, we just observed one peak which suggested no signs of introgression of *M. m. castaneus* into the KM mouse Chr 1.

We next performed sliding window similarity analysis using 500 kb windows and 100 kb sliding intervals (Figure 3(b)). We found that 13.5% and 6.4% of the Chr 1 sequences had high similarity (>99.7%) with the corresponding sequences of PWK and CAST, respectively. The distal portion of the B6-Chr1<sup>KM</sup> Chr 1 was found to have several regions that were highly similar to the corresponding regions of PWK with sharp boundaries between the regions of high and low similarity. However, we did not find any distinct boundaries between B6-Chr1<sup>KM</sup> and CAST Chr 1 sequence.

3.5. Bayesian Concordance Analysis. To determine the extent of phylogenetic discordance in B6-Chr1<sup>KM</sup> Chr 1, we assessed the discordance along Chr 1 by BCA. A total of 886 partitioned individual locus trees were used to estimate Bayesian concordance factors. In BCA, 87.7% of the loci supported a single KM/WSB topology with higher posterior probability, and 9.7% supported a single KM/PWK topology. None of the loci supported a KM/CAST topology, and the remaining 2.6% had a complicated topology (Figure 4(a)). Highly conserved genomic regions (Figure 3(b)) between the KM and PWK were almost found to have a relatively close topological relationship (Figure 4(a)). Furthermore, five loci with KM/WSB or KM/PWK topology were randomly

### TABLE 2: List of human disease-associated genes with loss of function variants in B6-Chr1<sup>KM</sup> Chr 1.

Algort syndrome, autosomal dominant     104200       Col4a3     ENSMUSG00000079465     Frameshift     Algort syndrome, autosomal recessive     203780       Hernaturia, bering familial, EPH     141200     Glomeralopathy with fibronectin deposits 2; GFND2     601894       Pales     ENSMUSG00000026239     Splice donor     Joubert syndrome, autosomal recessive     203780       Rab3gap2     ENSMUSG0000004709     Stop gair, splice donor     Rheatmatiod arthritis, RA     180300       Rab3gap2     ENSMUSG00000004799     Stop gair, splice donor     Macular degeneration, age-clated, 1; ARDD1     610425       Amelogenesis imperfects, type IA; AII A     104530     614225     Amelogenesis imperfects, type IA; AII A     104530       Lamb3     ENSMUSG0000002649     Missense     Xeroderma pigmentosum, complementation group G; XPG     226650       Dat     ENSMUSG0000002648     Missense     Xeroderma pigmentosum, complementation group G; XPG     674653       Ercc5     ENSMUSG00000026048     Missense     Veuropathy, hereditary sensory and autoomic, type V; HSAN     674653       Macus 200000026049     Missense     Xeroderma pigmentosum, complementation group G; XPG     275780       Caspp	Gene	Ensembl ID	Variant type	Phenotype	OMIM ID
Caléa ENSMUSG00000079465 Frameshift Alport syndrome, autosomal recessive (2000)   Fn1 ENSMUSG00000026193 Frameshift Clomerulopathy with fibronecti deficiency 61410   Pde6d ENSMUSC00000006203 Splice dono Joubert syndrome 22, IBTS22 6150   Cd244 ENSMUSG00000066424 Frameshift Macular degeneration, age-related, 1; RAMD1 60007   Cd244 ENSMUSG0000000000000000000000000 Splice acceptor, Martsoff syndrome 20270   Rab3gap ENSMUSG000000026639 Splice acceptor, Martsoff syndrome 20270   Rab3gap ENSMUSG00000026639 Splice acceptor, Martsoff syndrome 20270   Cappa ENSMUSG00000026648 Missense Keroderma ginementotional, Heritz type 22670   Enstructional, Receptor Missense Keroderma ginementotional autoonmic, type Vi, HA, ALL 20270   Cappa ENSMUSG00000026048 Missense Keroderma ginementotional group G, XPG 2771   Cappa ENSMUSG00000026017 Missense Leiderma ginementotion group G, XPG 27720   Cappa ENSMUSG0000002617 Missense Leiderbar syndrome 1; IBTS1 21320   Rest ENSMUSG0000002617 Missense Sectif Lappaga functional group G; KPG 27700   Distor <td></td> <td></td> <td></td> <td>Alport syndrome, autosomal dominant</td> <td>104200</td>				Alport syndrome, autosomal dominant	104200
Hermaturia, benign familial; BFH     14200       Fn1     ENSMUSC00000026193     Prameshift     Glomerulopathy with fibronectin deposits 2; GFND2     61804       Palesd     ENSMUSC00000002639     Splice donor     Joubert syndrome 22; JITS52     61805       Cd24     ENSMUSC00000000793     Stop gain; splice donor     Matcalar degeneration, age-related, 1; ARMD1     63005       Cd24     ENSMUSC000000039318     Splice acceptor, Matrixolf syndrome     212200       Rab3gap     ENSMUSC00000002663     Splice acceptor, Matrixolf syndrome 2; WARBM2     61625       LamM5     ENSMUSC0000002663     Splice acceptor, Epidermolysis bulloss, junctional, Herlitz type     22000       ENSMUSC0000002613     Missense     Neuropathy, hercitary sensory and autonomic, type VI, Britt     61625       Data     ENSMUSC0000002619     Missense     Keuropathy, hercitary sensory and autonomic, type VI, Britt     61605       Tame:17     ENSMUSC0000002619     Missense     Joubert syndrome; 13751     61303       Bardi     ENSMUSC0000002619     Missense     Fibriance Spritter, Stop Control     61303       Bardi     ENSMUSC0000002619     Missense     Specific language inpaintent 1; BTS14	Col4a3	ENSMUSG0000079465	Frameshift	Alport syndrome, autosomal recessive	203780
Final ENSMUSG000002013 Pramsshift Glomerulopathy with fibronectin deposits 2; GFND2 601894   Pdedd ENSMUSG000002239 Splice donor Julasma fibronectin deficiency 61101   Pdedd ENSMUSG0000006421 Frameshift Macular degeneration, age-related, 1; ARMD1 603075   Cd2/14 ENSMUSG0000006491 Stop gain; splice donor Rheumatoid arthritis; RA 10000   Rab3gap ENSMUSG0000002693 Splice acceptor, missense Warburg micro syndrome 2; WARM2 614225   Amatol Syndrome ENSMUSG000002663 Splice acceptor, Epidermolysis bullosa, junctional, non-Heritz type 26600   Dat ENSMUSG0000026131 Missense Neuropathy, hereditary sensory and autonomic, type VI; 614653   Dat ENSMUSG0000026048 Missense Xeroderma pigmentosum, complementation group G; XPG 27870   Caspa ENSMUSG000002609 Missense Joubert syndrome 1; JBTS1 21300   Tamen23 ENSMUSG000002617 Missense Joubert syndrome; IS 26000   Glade syndrome; IS Glades syndrome; IS 25000   Tamen23 ENSMUSG000002617 Missense Fraedes marterist, ISIS1 615425   Barsat ENSMUSG000002617 Missense Fraedes marterist, ISIS1 615425   Chigler-Najar syn				Hematuria, benign familial; BFH	141200
In1 ENSMUS.60000026193 FrameShift Plasma fibronectin deficiency 614101   Pde6d ENSMUS.600000026239 Splice donor Joubert syndrome 22; JBTS22 615665   Runcal ENSMUS.6000000040709 Stop gain; splice donor Rheumatoid arthritis; RA 180300   Rab3gap2 ENSMUS.600000002639 Splice acceptor, Martsol syndrome 22; WARM2 614225   Amelogenesis imperfecta, type 14, AIAIA 104330   Lamb3 ENSMUS.60000002643 Splice acceptor Epidermolysis bullosa, junctional, non Herlitz type 226500   Evidermolysis bullosa simpler, autosomal recessive 2; EBS2 615425 614525   Dst ENSMUS.60000002648 Missense Xeroderma pigmentosum, complementation group G; XPG 78780   Caspas ENSMUS.60000002649 Missense Joubert syndrome 1; JBTS1 21300   Tmem237 ENSMUS.60000002619 Missense Bread cancer 114480   Bard1 ENSMUS.60000002619 Missense Bread cancer 114480   Brad1 ENSMUS.600000026172 Missense Trace Mitchengarome 2; 3MZ 622200   Mistochondrial complex 11 deficiency, nuclear type 1; Missense Mitchendrial complex 1; BTS1 613425   Ds312 ENSMUS.600000026172 Missense Specific Inaguage impairment 5; SLIS 61543				Glomerulopathy with fibronectin deposits 2; GFND2	601894
Pde6dENSMUSG0000026239Splice donorJoubert syndrome 22; JETS2615665Hmcn1ENSMUSG00000060649FrameshiftMacular degeneration, age related, it ARD10603075Ca244ENSMUSG00000039318Splice acceptor, missenseMartsolf syndrome 2; WARBM218230Rab3gap2ENSMUSG00000026639Splice acceptor, missenseMartsolf syndrome 2; WARBM2262700Lamb3ENSMUSG00000026639Splice acceptorEpidermolysis bullosa, junctional, non-Heritz type26600DataENSMUSG000002648MissenseKeroderma pigmentosum, complementation group G; XPG77870Ca398ENSMUSG0000002609MissenseCASPase 8 deficiency603165Teme323ENSMUSG0000002609MissenseJoubert syndrome 1; JETS121300Tame324ENSMUSG0000002617MissenseJoubert syndrome 1; JETS121500Brad1ENSMUSG0000002617MissenseGracle syndrome, IS25000ChildENSMUSG000002617MissenseCaspla Specific language impairment, SLI25000MissenseSpecific language impairment, SLI2500025000ChildENSMUSG000002617MissenseFarme syndrome, IS25000ChildENSMUSG000002617MissenseSpecific language impairment, SLI25000MissenseSpecific language impairment, SLI2500025000MissenseSpecific language impairment, SLI25000ChildENSMUSG0000026139MissenseCrigler Najar syndrome, Lehd type, I26000Child	Fn1	ENSMUSG0000026193	Frameshift	Plasma fibronectin deficiency	614101
Hmen1ENSMUSG00000066842FrameshiftMacular degeneration, age-related, 1; ARMD1603075Cd244ENSMUSG0000000709Stop gain; splice donorRheunatoid arthritis; RA180200Rab3gap2ENSMUSG00000026639Splice acceptorMarteol syndrome 2; WARBM2614225AmabENSMUSG0000026639Splice acceptorEpidermolysis bullosa, junctional, non-Herlitz type226700EnsmusG0000026131MissenseKeuropathy, hereditary sensory and autonomic, type VI; HSAN6614635EnsmusG00000026048MissenseXeroderma pigmentosum, complementation group G; XPG2787800Casp8ENSMUSG00000026049MissenseXeroderma pigmentosum, complementation group G; XPG2787800Casp8ENSMUSG00000026029MissenseJoubert syndrome 1; JBTS1213300Tmem237ENSMUSG00000026196MissenseBreast cancer114482Biornatal syndrome; BJSCa2000Gracile syndrome 1; JBTS1213300ChissenseLeigh syndrome; 15256000MitsenseSperitic language inpairment 5; SLS615432Besl1ENSMUSG00000026172MissenseThree M syndrome; 15124000Obsl1ENSMUSG00000026179MissenseSperitic language inpairment 5; SLS615432MissenseSperitic language inpairment 5; SLS615432615432Dis312ENSMUSG00000026179MissenseSperitic language inpairment 5; SLS615432MissenseSperitic language inpairment 5; SLS615432615432Ug11aENSMUSG00000026193 </td <td>Pde6d</td> <td>ENSMUSG0000026239</td> <td>Splice donor</td> <td>Joubert syndrome 22; JBTS22</td> <td>615665</td>	Pde6d	ENSMUSG0000026239	Splice donor	Joubert syndrome 22; JBTS22	615665
Cd244ENSMUSG0000004709Stop gain: splice donor splice acceptor, missenseRheumatoi arthritis: RA180300Rab3gap2ENSMUSG0000039318Splice acceptor, missenseMarburg micro syndrome 2; WARBM2614225 Amelogenesis inperfecta, type IA; A11A104530Lamb3ENSMUSG0000026639Splice acceptor missenseEpidermolysis bullosa, junctional, herleitz type22650Ercc5ENSMUSG0000026131MissenseKaropathy, hereditary sensory and autoomic, type VI; HSAN6614525Casp8ENSMUSG0000026048MissenseXeroderma pigmentosum, complementation group G; XPG278780Casp8ENSMUSG00000026029MissenseLoubert syndrome 1; JBTS1213300Tmem237ENSMUSG00000026196MissenseJoubert syndrome 1; JBTS1213300Bes11ENSMUSG00000026196MissenseBiornstal syndrome; IS26000Gracile syndrome; IS256000MissenseGracile syndrome; IS256000MissenseSpecific language impairment 5; SL15615432Disl1ENSMUSG00000026112MissenseSpecific language impairment 5; SL15615432Disl2ENSMUSG00000026173MissenseSpecific language impairment 5; SL15615432Disl2ENSMUSG00000026179MissenseSpecific language impairment 5; SL15615432Disl2ENSMUSG00000026179MissenseSpecific language impairment 5; SL15615432Disl2ENSMUSG00000026179MissenseSpecific language impairment 5; SL15615432Disl2ENSMUSG00000026179	Hmcn1	ENSMUSG0000066842	Frameshift	Macular degeneration, age-related, 1; ARMD1	603075
Rab3gap2ENSMUSG00000039318Splice acceptor, missenseMartsolf syndrome 1212720 Marburg micro syndrome 2; WARDM2614235Lamb3ENSMUSG0000026639Splice acceptorEpidermolysis bullosa, junctional, non-Herlitz type226500Epidermolysis bullosa, junctional, non-Herlitz type226500Epidermolysis bullosa simplex, autosomal recessive 2; EBSE2615425DstENSMUSG0000026048MissenseXeroderma pigmentosum, complementation group G; XPG278780Casp8ENSMUSG00000026029MissenseDermatitis, atopic603165Tmem237ENSMUSG000002619MissenseJoubert syndrome 14; JBTS1614424Bard1ENSMUSG0000026196MissenseBreast cancer114480Bjornstad syndrome; IS226000Gracile syndrome; IS226000MitsenseBreast cancer114480BiorsEnsMUSG0000026172MissenseLeigh syndrome; IS226000Obs11ENSMUSG00000026172MissenseSpecific language impairment; S sU.5615432Dis32ENSMUSG00000026149MissenseSpecific language impairment; S sU.5615432Dis312ENSMUSG00000026253MissensePrirenan syndrome; PRLMNS267000Multiple pterygium syndrome; REMMUS267000Multiple pterygium syndrome; REMTS235900Oth1ENSMUSG00000026389MissensePrechman syndrome; PRLMNS267000Multiple pterygium syndrome; REMMUS265000Gibbert syndrome; REMTS235900Chiger-Najjar syndrome; REMMUS267000Gibbert sy	Cd244	ENSMUSG0000004709	Stop gain: splice donor	Rheumatoid arthritis: RA	180300
Rab3gap2   ENSMUSG0000003318   missense   Warburg micro syndrome 2; WARBM2   614225     Lamb3   ENSMUSG0000026639   Splice acceptor   Epidermolysis bullosa, junctional, Heritiz type   226700     Epidermolysis bullosa, junctional, non-Heritiz type   226700   Epidermolysis bullosa, junctional, non-Heritiz type   226700     Dst   ENSMUSG000000260131   Missense   Xeroderma pigmentosum, complementation group G; XPG   278780     Casp8   ENSMUSG00000026029   Missense   Xeroderma pigmentosum, complementation group G; XPG   278780     Casp8   ENSMUSG00000026196   Missense   Joubert syndrome 1; JBTS1   213300     Casp1   ENSMUSG00000026196   Missense   Joubert syndrome 1; JBTS1   213300     Bard1   ENSMUSG00000026172   Missense   Breast cancer   114480     Bjort acceptor   Missense   Three M syndrome; JS   256000     Mitochondrial complex III deficiency, nuclear type 1;   124000   124000     Obs11   ENSMUSG0000002611   Missense   Specific language impairment 5; SI15   615432     Dis312   ENSMUSG00000026121   Missense   Perlman syndrome; PRLIMNS   265000     Ug11a1   <	04211		Splice acceptor, missense	Martsolf syndrome	212720
	Rab3gap2	ENSMUSG0000039318		Warburg micro syndrome 2: WARBM2	614225
Lamb3 ENSMUSG0000026639 Splice acceptor Epidermolysis bullosa, junctional, Herlitz type 226500 Epidermolysis bullosa, junctional, Herlitz type 226500 CASPase 8 deficiency 607271 Caspa 8 ENSMUSG00000026029 Missense Joubert syndrome 1; JBTS14 614424 Bard1 ENSMUSG00000026196 Missense Joubert syndrome 1; JBTS14 614424 Bard1 ENSMUSG00000026172 Missense Leigh syndrome; BJS 265000 Mitochondrial complex III deficiency, nuclear type 1; 124000 Mitochondrial complex III deficiency, nuclear type 1; 124000 Mitole type of the syndrome; 5, SLI5 615432 ENSMUSG00000026211 Missense Specific language impairment 5; SLI5 615432 ENSMUSG0000002623 Missense Specific language impairment 5; SLI5 615432 ENSMUSG0000002623 Missense Perlman syndrome; PRLMNS 267000 Multiple pterygium syndrome, tehal type; LMPS 253290 Crigler-Najjar syndrome, type I 218800 Crigler-Najjar syndrome, type I 21800 Multiple pterygium syndrome, type I 21800 Hyperbilirubinemia, transitient familial neonatal; HBLRTFN 253290 Keap3 ENSMUSG0000026429 Missense Fanconi anemia, complementation group T; FANCT 616435 Pox ENSMUSG0000002632 Missense				Amelogenesis imperfecta, type IA: AIIA	104530
$ \begin{array}{c} \label{eq:harmonic} \begin{tabular}{lllllllllllllllllllllllllllllllllll$	Lamb3	ENSMUSG0000026639	Splice acceptor	Epidermolysis bullosa, junctional. Herlitz type	226700
Disk     ENSMUSG0000026131     Missense     Epidermolysis bullosa inplex, autosonal necessive 2; EBSB2     615425       Dst     ENSMUSG0000026048     Missense     Neuropathy, hereditary sensory and autonomic, type VI; HSAN6     614653       Ercc5     ENSMUSG0000026029     Missense     Xeroderma pigmentosum, complementation group G; XPG     278780       Casp8     ENSMUSG0000026029     Missense     Dermatitis, atopic     607271       Tmem237     ENSMUSG0000026026     Missense     Dermatitis, atopic     603165       Joubert syndrome 14; JBTS14     614424     614424     614424     614424     614424       Bard1     ENSMUSG0000026172     Missense     Breast cancer     114480       Bjornstad syndrome; IJS     256000     603358     256000       Obsl1     ENSMUSG0000002611     Missense     Three M syndrome; IS     250000       Obsl2     ENSMUSG00000026149     Missense     Specific language impairment 5; SLI5     615432       Dis312     ENSMUSG00000026333     Missense     Perlman syndrome, Hethal type; LMPS     255200       Chrigh Proxylian syndrome, type II     606785     Gilder- syndrom	Lunice	111011100000000000000000000000000000000		Epidermolysis bullosa junctional non-Herlitz type	226650
Dst   ENSMUSG0000026131   Missense   Neuropathy, heroditry sensory and autonomic, type VI; HSAN6   614653     Ercc5   ENSMUSG0000026048   Missense   Xeroderma pigmentosum, complementation group G; XPG   278780     Casp8   ENSMUSG0000026029   Missense   Dermatitis, atopic   603165     Tmem237   ENSMUSG00000026196   Missense   Joubert syndrome 1; JBTS1   213300     Bard1   ENSMUSG00000026196   Missense   Breast cancer   114480     Biornstad syndrome; BJS   262000   Gracile syndrome   603358     Bcs11   ENSMUSG00000026172   Missense   Breast cancer   114480     Obs11   ENSMUSG00000026172   Missense   Specific language impairment 5; SLI5   615432     Tm4s20   ENSMUSG00000026149   Missense   Specific language impairment 5; SLI5   615432     Dis312   ENSMUSG00000026233   Missense   Perlman syndrome, PRIMNS   265000     Chrigg   ENSMUSG00000026238   Missense   Crigler-Najjar syndrome, type 1   218800     Clapt-Najjar syndrome, ISMUSG00000026238   Missense   Crigler-Najjar syndrome, type 1   218800     Ugt1a1   ENSMUSG00000026429				Endermolysis bullosa simpley autosomal recessive 2: FBSB2	615425
Ercc5   ENSMUSG0000026048   Missense   Xeroderma pigmentosum, complementation group G; XPG   278780     Casp8   ENSMUSG00000026029   Missense   CASPase 8 deficiency   603165     Tmem237   ENSMUSG00000026019   Missense   Joubert syndrome 1; JBTS1   213300     Joubert syndrome 14; JBTS14   614424   614424     Bard1   ENSMUSG00000026196   Missense   Breast cancer   114480     Bard1   ENSMUSG00000026172   Missense   Leigh syndrome; BJS   226000     Gracile syndrome   Gracile syndrome; CS   256000   Mitochondrial complex III deficiency, nuclear type 1; MG3DN1   124000     Obs11   ENSMUSG00000026211   Missense   Specific language impairment 5; SL15   615432     Dis312   ENSMUSG00000026214   Missense   Pertman syndrome; PRLMNS   267000     Multiple pterygium syndrome, is yndrome, is sufficiency, nuclear type 1; Crigler-Najjar syndrome, type 1   218800   21890     Chrmg   ENSMUSG00000026253   Missense   Pertman syndrome; type 1   218900     Crigler-Najjar syndrome, type I   1066785   Gibert syndrome   143500     Ugt1a1   ENSMUSG00000026389   Missense	Dst	ENSMUSG0000026131	Missense	Neuropathy hereditary sensory and autonomic type VI-	013423
Ercc5ENSMUSG0000026048MissenseXeroderma pigmentosum, complementation group G; XPG278780Casp8ENSMUSG0000026029MissenseCASPase 8 deficiency607271Dermatifis, atopic603165Tmem237ENSMUSG00000038079MissenseJoubert syndrome 1; JBTS1213300Joubert syndrome 14; JBTS14614424Bard1ENSMUSG0000026196MissenseBreast cancer114480Bjornstad syndrome; BJS262000Gracile syndrome263000Bes11ENSMUSG0000026172MissenseLeigh syndrome; LS256000Mitochondrial complex III deficiency, nuclear type 1; MG3D1112400012000Obs11ENSMUSG0000026211MissenseThree M syndrome; S124000Obs12ENSMUSG0000026149MissenseSpecific language impairment 5; SL15615432Dis312ENSMUSG00000026149MissensePerlman syndrome, tPKLMNS267000Chrigger-Najjar syndrome, tPKLMNS267000Multiple pterygium syndrome, tscobar variant; EVMPS253290Ugt1a1ENSMUSG00000026253MissensePerlman syndrome, type 1218800Ugt1a1ENSMUSG00000026389MissenseFanconi anemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG00000026372MissenseFanconi anemia, complementating group T; FANCT61635PpoxENSMUSG00000026729MissenseFanconi anemia, complementating group T; FANCT616435PpoxENSMUSG00000026372MissenseFanconi anemia, complementating group T; FANCT <td>200</td> <td></td> <td>Wilsselise</td> <td>HSAN6</td> <td>614653</td>	200		Wilsselise	HSAN6	614653
Casp8ENSMUSG0000026029MissenseCASPase 8 deficiency607271Tmem237ENSMUSG00000038079MissenseDermatitis, atopic603165Tmem237ENSMUSG00000026196MissenseJoubert syndrome 1; JBTS121300Bard1ENSMUSG0000026196MissenseBreast cancer114480Bard1ENSMUSG0000026172MissenseBreast cancer114480Bcs11ENSMUSG0000026172MissenseLeigh syndrome; IS256000Mitochondrial complex III deficiency, nuclear type 1; MC3DN124000615232Obs11ENSMUSG0000026211MissenseSpecific language impairment 5; SLI5615432Tm4520ENSMUSG0000026214MissenseSpecific language impairment 5; SLI5615432Dis312ENSMUSG0000026233MissensePerlman syndrome; PRLMNS267000Chrigter-Najjar syndrome, type I218800Crigter-Najjar syndrome, type I218800Ugt1a1ENSMUSG0000026389MissenseCrigter-Najjar syndrome, type I606785Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Anemia, hypochromic microytic, with iron overload 2; AHMIO2615234Ubc2tENSMUSG0000026389MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000026372MissenseParconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000026372MissenseParconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000037872MissenseParconi anemia, co	Ercc5	ENSMUSG0000026048	Missense	Xeroderma pigmentosum, complementation group G; XPG	278780
CaspsDermatitis, atopic603165Tmem237ENSMUSG0000038079MissenseJoubert syndrome 14; JBTS1213300Bard1ENSMUSG0000026196MissenseBreast cancer114480Bard1ENSMUSG0000026196MissenseBreast cancer114480Bard1ENSMUSG0000026172MissenseLeigh syndrome; BJS262000Gracile syndrome;ENSMUSG0000026172MissenseLeigh syndrome; DS262000Obs11ENSMUSG0000026211MissenseSpecific language impairment 5; SL15615432Dis32ENSMUSG0000026214MissenseSpecific language impairment 5; SL15615432Dis32ENSMUSG0000026253MissensePerlman syndrome; PRLMNS267000ChrngENSMUSG00000026253MissenseCrigler-Najjar syndrome, type I218800Chrigt+ENSMUSG00000026389MissenseCrigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG00000026389MissenseFanconi anemia, complementation group T; FANCT616435PopoxENSMUSG00000026389MissenseFanconi anemia, complementation group T; FANCT616435PopoxENSMUSG00000026329MissensePanconi anemia, complementation group T; FANCT616435Spherocytosis 2; EL2130600Spherocytosis 2; EL2130600	Com	ENSMUSCOOOOO26020	Missonso	CASPase 8 deficiency	607271
Immem 237ENSMUSG0000038079MissenseJoubert syndrome 1; JBTS1213300Bard1ENSMUSG0000026196MissenseBreast cancer114480Bard1ENSMUSG0000026192MissenseBjornstad syndrome; BJS26000Biornstad syndrome; LS256000Gracile syndrome; LS256000Bard1ENSMUSG0000026212MissenseLeigh syndrome; LS26000Obs11ENSMUSG0000026211MissenseThree M syndrome; 2; 3 M2612921Tm4s20ENSMUSG000002619MissenseSpecific language impairment 5; SLI5615432Dis312ENSMUSG000002619MissensePerlman syndrome; PRLMNS267000ChrngENSMUSG0000026253MissenseMultiple pterygium syndrome, texobar variant; EVMPS253290Ugr1a1ENSMUSG0000026639MissenseCrigler-Najjar syndrome, type I206702Ub224ENSMUSG000002629MissenseAnemia, hypochromic incrocytic, with iron overload 2; AHMID2615234Ub241ENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PipoxENSMUSG0000026429MissenseMalaria, succeptibility to611162PipoxENSMUSG0000026532MissensePropoynia variegata176200Steap3ENSMUSG0000026532MissenseProphyria variegata176200Spherocytosis, type 3; SPH3270970270970270970Steap4ENSMUSG0000026532MissensePyropoilalocytosis, type 3; SPH3270970Spherocytosis, type 3; SPH3 <td< td=""><td>Casp8 ENSMUSG0000026025</td><td>EIN310103G00000020029</td><td>Missense</td><td>Dermatitis, atopic</td><td>603165</td></td<>	Casp8 ENSMUSG0000026025	EIN310103G00000020029	Missense	Dermatitis, atopic	603165
Intent237   ENSMUSG000000263079   Missense   Joubert syndrome 14; JBTS14   614424     Bard1   ENSMUSG0000026196   Missense   Breast cancer   114480     Bjornstad syndrome; BJS   262000   Gracile syndrome   603358     Bcs11   ENSMUSG00000026172   Missense   Leigh syndrome; LS   256000     Obs11   ENSMUSG00000026171   Missense   Three M syndrome; 2; 3 M2   612921     Tm4s120   ENSMUSG00000026149   Missense   Specific language impairment 5; SL15   615432     Dis312   ENSMUSG00000026149   Missense   Perlman syndrome; PRLMNS   265000     Chrug   ENSMUSG00000026253   Missense   Perlman syndrome, lethal type; LMPS   265000     Chrug   ENSMUSG00000026253   Missense   Multiple pterygium syndrome, lethal type; LMPS   253290     Crigler-Najjar syndrome   Crigler-Najjar syndrome, type I   218800   237900     Steap3   ENSMUSG00000026429   Missense   Anemia, hypochromic microcytic, with iron overload 2; AHMID2   615234     Ub22t   ENSMUSG00000026429   Missense   Fanconi anemia, complementation group T; FANCT   616435     Ppox   ENSMUSG0000002	Tm	ENGMERCOOOOO220070		Joubert syndrome 1; JBTS1	213300
Bard1   ENSMUSG0000026196   Missense   Breast cancer   114480     Bjornstad syndrome; BJS   262000     Bcs11   ENSMUSG0000026172   Missense   Cracile syndrome   256000     Mitochondrial complex III deficiency, nuclear type 1;   124000     Obs11   ENSMUSG0000026112   Missense   Three M syndrome 2; 3 M2   612921     Tm4sf20   ENSMUSG0000026149   Missense   Specific language impairment 5; SLI5   615432     Dis312   ENSMUSG0000026253   Missense   Perlman syndrome; PRLMNS   265000     Chrigter-Najjar syndrome, lethal type; LMPS   253200   Crigter-Najjar syndrome, type 1   218800     Qut1a1   ENSMUSG0000026253   Missense   Crigter-Najjar syndrome, type 1   218800     Ugt1a1   ENSMUSG00000262639   Missense   Crigter-Najjar syndrome, type 1   237900     Steap3   ENSMUSG0000026429   Missense   Fanconi anemia, hopochromic microcytic, with iron overload 2;   143500     Ube2t   ENSMUSG0000026729   Missense   Fanconi anemia, complementation group T; FANCT   6164351     Ppox   ENSMUSG00000026729   Missense   Porphyria variegata   176200     <	1 mem257	ENSM03G000000380/9	Missense	Joubert syndrome 14; JBTS14	614424
Biornstad syndrome; BJS 262000   Gracile syndrome; 603358   Bcs1l ENSMUSG0000026172 Missense Leigh syndrome; LS 256000   Mitochondrial complex III deficiency, nuclear type 1; MC3DN1 24000   Obs11 ENSMUSG0000026211 Missense Three M syndrome 2; 3 M2 612921   Tm4sf20 ENSMUSG0000026149 Missense Specific language impairment 5; SLI5 615432   Dis32 ENSMUSG0000026253 Missense Perlman syndrome; PRLMMS 265000   Chrug ENSMUSG0000026253 Missense Multiple ptergyium syndrome, tescobar variant; EVMPS 253000   Ug1a1 ENSMUSG00000026389 Missense Crigler-Najjar syndrome, type I 218800   Ug1a1 ENSMUSG0000026389 Missense Anemia, hypochronic microcytic, with iron overload 2; Alfmetink, hypochronic microcytic, with iron overload 2; Alfmetink 615234   Ub210 ENSMUSG0000026389 Missense Fanconi anemia, complementation group T; FANCT 616435   Pipox ENSMUSG0000026429 Missense Porphyria variegata 176200   Ackri1 ENSMUSG0000026729 Missense Porphyria variegata 176200   Ackri2 ENSMUSG00000037872 Missense Pyropoikliocytosis, hereditar;; HPP 261401   Spherocytosis, hyre 3;	Bard1	ENSMUSG0000026196	Missense	Breast cancer	114480
BeshENSMUSG0000026172MissenseGracile syndrome; LS603358BeshENSMUSG0000026211MissenseLeigh syndrome; LS256000Mitochondrial complex III deficiency, nuclear type 1; MG3DN1124000ObslENSMUSG0000026211MissenseSpecific language impairment 5; SLI5615432Dis3l2ENSMUSG00000026149MissensePerlman syndrome; PRLMNS267000ChrmgENSMUSG00000026253MissenseMultiple pterygium syndrome, Escobar variant; EVMPS253290Crigler-Najjar syndrome, type I218800Ug1a1ENSMUSG0000026389MissenseCrigler-Najjar syndrome, type I287900Steap3ENSMUSG0000026429MissenseGibert syndrome143500Ub2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG000002632MissenseFanconi anemia, complementation group T; FANCT616436PpoxENSMUSG0000026532MissensePorphyria variegata176200AckriENSMUSG0000026532MissensePyropoidilocytosis, hereditary; HPP266140Spherocytosis, sp.et, SPH3270970Epoxide hydrolase 1, microsonal; EPHX132810EpixelENSMUSG0000038776MissensePyropoidilocytosis, hereditary; HPP266140EpixelENSMUSG0000038776MissensePyropoide hydrolase 1, microsonal; EPHX132810EpixelENSMUSG0000038776MissensePyreclampsia/clampsia 1; PEE1198800				Bjornstad syndrome; BJS	262000
Bcs1l   ENSMUSG0000026172   Missense   Leigh syndrome; LS   256000     Mitochondrial complex III deficiency, nuclear type 1; MC3DN1   124000     Obs1l   ENSMUSG0000026211   Missense   Three M syndrome 2; 3 M2   612921     Tm4sf20   ENSMUSG0000026149   Missense   Specific language impairment 5; SLI5   615432     Dis3l2   ENSMUSG0000026253   Missense   Perlman syndrome; PRLMNS   265000     Chrng   ENSMUSG0000026253   Missense   Multiple pterygium syndrome, Escobar variant; EVMPS   253290     Crigler-Najjar syndrome, type I   218800     Ugt1a1   ENSMUSG0000026289   Missense   Crigler-Najjar syndrome, type II   606785     Ityperbilirubinemia, transient familial neonatal; HBLRTFN   237900   149200     Steap3   ENSMUSG0000026389   Missense   Anemia, hypochromic microcytic, with iron overload 2; AHMI02   615234     Ube2t   ENSMUSG00000026429   Missense   Paconi anemia, complementation group T; FANCT   616435     Ppox   ENSMUSG00000037872   Missense   Porphyria variegata   176200     Ackr1   ENSMUSG00000026532   Missense   Pyropoikilocytosis, hereditary; HPP   266140				Gracile syndrome	603358
$ \begin{array}{c ccc} \mbox{Mitochondrial complex III deficiency, nuclear type 1; $MC3DN1} & 124000 \\ \mbox{MC3DN1} & ENSMUSG0000026211 & Missense & Three M syndrome 2; 3 M2 & 612921 \\ Tm4sf20 & ENSMUSG0000026149 & Missense & Specific language impairment 5; SL15 & 615432 \\ Dis3l2 & ENSMUSG00000053333 & Missense & Perlman syndrome; PRLMNS & 267000 \\ \mbox{Chrng} & ENSMUSG0000026253 & Missense & Multiple pterygium syndrome, Escobar variant; EVMPS & 253290 \\ \mbox{Crigler-Najjar syndrome, type I} & 218800 \\ \mbox{Crigler-Najjar syndrome, type I} & 606785 \\ \mbox{Gilbert syndrome} & Missense & Crigler-Najjar syndrome, type I & 606785 \\ \mbox{Gilbert syndrome} & Missense & Crigler-Najjar syndrome, type I & 237900 \\ \mbox{Steap3} & ENSMUSG0000026289 & Missense & Anemia, hypochromic microcytic, with iron overload 2; \\ \mbox{Popx} & ENSMUSG0000026429 & Missense & Fanconi anemia, complementation group T; FANCT & 616435 \\ \mbox{Popx} & ENSMUSG00000026429 & Missense & Porphyria variegata & 176200 \\ \mbox{Ackr1} & ENSMUSG00000026729 & Missense & Porphyria variegata & 176200 \\ \mbox{Steap3} & ENSMUSG00000026532 & Missense & Pyropoikilocytosis, hereditary; HPP & 266140 \\ \mbox{Splar} & ENSMUSG00000026532 & Missense & Pyropoikilocytosis, hereditary; HPP & 266140 \\ \mbox{Splar} & ENSMUSG00000026532 & Missense & Pyropoikilocytosis, hereditary; HPP & 266140 \\ \mbox{Splar} & ENSMUSG00000037876 & Missense & Pyropoikilocytosis, hereditary; HPP & 266140 \\ \mbox{Spherocytosis, type 3; SPH3 & 270970 \\ \mbox{Epxide hydrolase 1, microsonal; EPHX1 & 132810 \\ \mbox{Epxide hydrolase 1, microsonal; EPHX1 & 132810 \\ \mbox{Epxide hydrolase 1, microsonal; IPEE1 & 189800 \\ \mbox{Chrosonal} & Preeclampsia/clampsia 1; PEE1 & 189800 \\ \mbox{Chrosonal} & Preclampsia/clampsia 1; PEE1 & 189800 \\ \mbox{Chrosonal} & Preclampsia/cl$	Bcs1l	ENSMUSG0000026172	Missense	Leigh syndrome; LS	256000
ObsliENSMUSG0000026211MissenseThree M syndrome 2; 3 M2612921Tm4sf20ENSMUSG0000026149MissenseSpecific language impairment 5; SLI5615432Dis3l2ENSMUSG0000053333MissensePerlman syndrome; PRLMNS267000ChrngENSMUSG0000026253MissensePerlman syndrome; Escobar variant; EVMPS253290Multiple pterygium syndrome, lethal type; LMPS253290Crigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG0000089960MissenseCrigler-Najjar syndrome, type II606785BuSMUSG0000026253MissenseCrigler-Najjar syndrome, type II605785Ugt1a2ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000026729MissenseMalaria, susceptibility to611162Spata1ENSMUSG0000026332MissensePyropoikilocytosis, hereditary; HPP266140Spata1ENSMUSG0000026532MissensePyropoikilocytosis, hereditary; HPP266140Epoxide hydrolase 1, microsonal; EPHX1132810Epoxide hydrolase 1, mi				Mitochondrial complex III deficiency, nuclear type 1; MC3DN1	124000
Tm4sf20ENSMUSG0000026149MissenseSpecific language impairment 5; SLI5615432Dis3l2ENSMUSG0000053333MissensePerlman syndrome; PRLMNS267000ChrngENSMUSG0000026253MissenseMultiple pterygium syndrome; Escobar variant; EVMPS253290Ugt1a1ENSMUSG00000089960MissenseCrigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG0000026389MissenseCrigler-Najjar syndrome, type II606785Gilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000026332MissensePorphyria variegata176200Ackr1ENSMUSG0000002632MissensePorpoikilocytosis, hereditary; HPP266140Spla1ENSMUSG00000026332MissensePyropoikilocytosis, hereditary; HPP266140Spla1ENSMUSG00000026332MissensePyropoikilocytosis, hereditary; HPP266140Ephx1ENSMUSG0000026332MissensePyropoikilocytosis, hereditary; HPP266140Spla1ENSMUSG0000038776MissensePyropoikilocytosis, hereditary; HPP266140Ephx1aENSMUSG0000038776MissensePyropoikilocytosis, hereditary; HPP266140Preeclampsia/cclampsia 1; PEE1ENSMUSG0000038776MissensePyropoikilocytosis, hereditary; HPP266140<	Obsl1	ENSMUSG0000026211	Missense	Three M syndrome 2; 3 M2	612921
Dis3l2ENSMUSG0000053333MissensePerlman syndrome; PRLMNS267000ChrngENSMUSG0000026253MissenseMultiple pterygium syndrome; Escobar variant; EVMPS253290Ugt1a1ENSMUSG00000089960MissenseCrigler-Najjar syndrome, lethal type; LMPS253290Ugt1a1ENSMUSG00000026389MissenseCrigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG00000026389MissenseGilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN23790037900Steap3ENSMUSG0000026389MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000026429MissenseFanconi anemia, susceptibility to611162ChristianENSMUSG00000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Epoxide hydrolase 1, microsomal; EPHX1132810Ephx1ENSMUSG0000038776MissenseHypercholanemia, familial; FHCA607748Preclampsia/cclampsia 1; PEE1189800	Tm4sf20	ENSMUSG0000026149	Missense	Specific language impairment 5; SLI5	615432
ChrngENSMUSG0000026253MissenseMultiple pterygium syndrome, Escobar variant; EVMPS265000Ugt1a1ENSMUSG0000089960MissenseCrigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG0000026389MissenseCrigler-Najjar syndrome, type II606785Gilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG00000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000037872MissensePorphyria variegata176200Ackr1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spta1ENSMUSG00000038776MissensePyropoikilocytosis, hereditary; HPP266140Ephx1ENSMUSG0000038776MissensePyropoikilocytosis, hereditary; HPA132810Ephx1ENSMUSG0000038776MissenseHypercholanemia, familial; FHCA60748Preeclampsia/eclampsia 1; PEE1189800	Dis3l2	ENSMUSG0000053333	Missense	Perlman syndrome; PRLMNS	267000
Chrng ENSMUSG00000026253 Missense Multiple pterygium syndrome, lethal type; LMPS 253290 Crigler-Najjar syndrome, type I 218800 Crigler-Najjar syndrome, type II 606785 Gilbert syndrome 143500 Hyperbilirubinemia, transient familial neonatal; HBLRTFN 237900 Steap3 ENSMUSG0000026389 Missense Anemia, hypochromic microcytic, with iron overload 2; AHMIO2 615234 Ube2t ENSMUSG00000026429 Missense Fanconi anemia, complementation group T; FANCT 616435 Ppox ENSMUSG00000062729 Missense Porphyria variegata 176200 Ackr1 ENSMUSG00000037872 Missense Malaria, susceptibility to 611162 Elliptocytosis 2; EL2 130600 Spta1 ENSMUSG00000026532 Missense Pyropoikilocytosis, hereditary; HPP 266140 Spherocytosis, type 3; SPH3 270970 Epoxide hydrolase 1, microsomal; EPHX1 132810 Ephx1 ENSMUSG00000038776 Missense Hypercholanemia, familial; FHCA 607748 Preeclampsia/eclampsia 1; PEE1 189800			Multiple ptervgium syndrome, Escobar variant; EVMPS	265000	
Ugt1a1ENSMUSG0000089960MissenseCrigler-Najjar syndrome, type I218800Ugt1a1ENSMUSG00000089960MissenseCrigler-Najjar syndrome, type II606785Gilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG00000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162Spta1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Chrng	ENSMUSG0000026253	Missense	Multiple pterygium syndrome, lethal type; LMPS	253290
Ugt1a1ENSMUSG0000089960MissenseCrigler-Najjar syndrome, type II606785Gilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162Spta1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spharocytosis, type 3; SPH3270970Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800				Crigler-Najjar syndrome, type I	218800
Ugt1a1ENSMUSG00000089960MissenseGilbert syndrome143500Gilbert syndromeGilbert syndrome143500Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG00000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162Elliptocytosis 2; EL2130600Spta1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800				Crigler-Najjar syndrome, type II	606785
Hyperbilirubinemia, transient familial neonatal; HBLRTFN237900Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG00000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162ENSMUSG0000026532MissensePyropoikilocytosis 2; EL2130600Spta1ENSMUSG0000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG0000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Ugt1a1	ENSMUSG0000089960	Missense	Gilbert syndrome	143500
Steap3ENSMUSG0000026389MissenseAnemia, hypochromic microcytic, with iron overload 2; AHMIO2615234Ube2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spta1ENSMUSG00000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800				Hyperbilirubinemia, transient familial neonatal; HBLRTFN	237900
Ube2tENSMUSG0000026429MissenseFanconi anemia, complementation group T; FANCT616435PpoxENSMUSG0000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162Elliptocytosis 2; EL2130600Spta1ENSMUSG0000026532MissensePyropoikilocytosis, hereditary; HPP266140Spta1ENSMUSG00000026532MissensePyropoikilocytosis, type 3; SPH3270970Epoxide hydrolase 1, microsomal; EPHX1132810Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Steap3	ENSMUSG0000026389	Missense	Anemia, hypochromic microcytic, with iron overload 2;	615234
PpoxENSMUSG0000062729MissensePorphyria variegata176200Ackr1ENSMUSG00000037872MissenseMalaria, susceptibility to611162Elliptocytosis 2; EL2130600Spta1ENSMUSG0000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Ube2t	ENSMUSG0000026429	Missense	Fanconi anemia, complementation group T; FANCT	616435
Ackr1   ENSMUSG00000037872   Missense   Malaria, susceptibility to   611162     Elliptocytosis 2; EL2   130600     Spta1   ENSMUSG00000026532   Missense   Pyropoikilocytosis, hereditary; HPP   266140     Spherocytosis, type 3; SPH3   270970     Ephx1   ENSMUSG00000038776   Missense   Hypercholanemia, familial; FHCA   607748     Preeclampsia/eclampsia 1; PEE1   189800	Ppox	ENSMUSG0000062729	Missense	Porphyria variegata	176200
Spta1ENSMUSG0000026532MissenseFiliptocytosis 2; EL2130600Spta1ENSMUSG0000026532MissensePyropoikilocytosis, hereditary; HPP266140Spherocytosis, type 3; SPH3270970Ephx1ENSMUSG0000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Ackr1	ENSMUSG0000037872	Missense	Malaria, susceptibility to	611162
Spta1   ENSMUSG0000026532   Missense   Pyropoikilocytosis, hereditary; HPP   266140     Spharocytosis, type 3; SPH3   270970     Ephx1   ENSMUSG0000038776   Missense   Hypercholanemia, familial; FHCA   607748     Preeclampsia/eclampsia 1; PEE1   189800				Elliptocytosis 2: EL2	130600
EpixalEnvironmentEnvironmentEnvironmentEnvironmentSpherocytosis, type 3; SPH3270970EpixalEpoxide hydrolase 1, microsomal; EPHX1132810EpixalENSMUSG00000038776MissenseHypercholanemia, familial; FHCA607748Preeclampsia/eclampsia 1; PEE1189800	Spta1	ENSMUSG0000026532	Missense	Pyropoikilocytosis, hereditary: HPP	266140
Ephx1   ENSMUSG00000038776   Missense   Epoxide hydrolase 1, microsomal; EPHX1   132810     Preeclampsia/eclampsia 1; PEE1   189800	-F.m.1		1110001100	Spherocytosis type 3: SPH3	270970
Ephx1   ENSMUSG00000038776   Missense   Hypercholanemia, familial; FHCA   607748     Preeclampsia/eclampsia 1; PEE1   189800				Epoxide hydrolase 1 microsomal: FPHX1	132810
Preeclampsia/eclampsia 1; PEE1 189800	Ephy1	ENSMUSG0000038776	Missense	Hypercholanemia familial: FHCA	607748
Trectampsia Clampsia 1, TEET 109000	-P	21.011000000000000000000000000000000000	1110501150	Preeclampsia/eclampsia 1. PFF1	189800
Rd3 ENSMUSG00000049353 Missense Leber concentral amaurosis 12: I CA12 610612	Rd3	ENSMUSG0000049353	Missense	Leber congenital amaurosis 12. ICA12	610612

TABLE 2: Continued.

Gene	Ensembl ID	Variant type	Phenotype	OMIM ID
Cd46	ENSMUSG0000016493	Missense	Hemolytic uremic syndrome, atypical, susceptibility to, 2; AHUS2	612922
			Immunodeficiency, common variable, 2; CVID2	240500
Cr2	ENSMUSG00000026616	Missense	Immunodeficiency, common variable, 7; CVID7	614699
			Systemic lupus erythematosus, susceptibility to, 9; SLEB9	610927

OMIM: online Mendelian inheritance in man. Numbers in italic in OMIM ID column indicate that these diseases have mouse models. Human diseaserelated phenotypes come from "Human-Mouse: Disease Connection" database (http://www.informatics.jax.org/humanDisease.shtml) in Mouse Genome Informatics website.

selected, and the phylogenetic trees were confirmed by Mega software (Figure 4(b)).

#### 4. Discussion

Because the KM mouse is used regularly in pharmaceutical and genetic studies, its detailed genetic structure is of great value to the research community. In this study, we sequenced the genome of a male B6-Chr1<sup>KM</sup> mouse, in which Chr 1 was derived from one KM mouse. The detailed sequence analysis would provide new insights into the application of B6-Chr1<sup>KM</sup> in biomedical research.

In this study, we identified 479,956 SNPs and 96,679 indels on Chr 1, of which 8.1% did not exist in the MGP and dbSNP142 data sets, indicating that these variants were unique to the B6-Chr1<sup>KM</sup> mice. Therefore, these variants can be used as unique genetic markers for the genetic quality control of KM mouse. As the most common types of genetic variants, SNPs and indels have been increasingly recognized as having a wide range of effects on gene functions. Among the variants identified on Chr 1, most were located within intergenic or intronic regions. However, we also identified 474 functional variants (missense variant with SIFT < 0.05, stop gain or loss variant, frameshift variant, and splice donor or acceptor variant) which influenced 155 genes. Additionally, several genes have been identified to be associated with human diseases, making them interesting candidates for further functional studies using KM mouse or our newly build B6-Chr1KM strain. For example, Rd3, which is associated with retinal degeneration, was identified as a missense substitution (A->T) with significant deleterious effects (p = 0.02). Previous studies have shown that mice with a homozygous mutation in Rd3 exhibit retinal degeneration at three weeks after birth [15]. We also identified a splice acceptor variant in Lamb3 gene, which is associated with blistering of the skin. The mouse models with homozygous Lamb3 628 G->A showed blistering and erosions after birth [16].

Since KM mouse is originated from Swiss mice, it has been speculated to be contaminated with *M. m. castaneus*. In 1991, the morphological characteristics and isozyme polymorphisms of KM and Swiss mice were evaluated, revealing the presence of distinct genetic differences between them [17]. Comparison of KM mouse with wild mice of *M. m. castaneus* captured in Kunming has revealed that the former is more closely related to *M. m. domesticus* than to *M. m. castaneus*. Conversely, contamination of KM mouse

by M. m. castaneus has been previously demonstrated using the isozyme test [18]. In 2003, the results of a study involving the detection of isozyme polymorphisms also supported the grouping of KM and Swiss mice with M. m. domesticus and not with M. m. musculus or M. m. castaneus [2]. However, it has not yet been confirmed whether KM mouse contains part of the genome of M. m. musculus or M. m. castaneus. Therefore, high resolution studies of Chr 1 of KM mouse by next-generation sequencing may clarify whether these mice were originated from Swiss mice and/or other mice. Our sequence similarity analysis provided substantial evidence that KM mouse was derived from *M. m. domesticus*, which means that Swiss mice were their ancestor. Both 100 kb blocks and sliding window similarity analysis demonstrated that the Chr 1 of KM mouse was largely composed of M. m. domesticus sequences with the rest may derive from M. m. musculus or M. m. castaneus. Therefore, further analysis is needed to determine the proportion of each subspecies contribution to the Chr 1 of KM mouse.

With the increasing number of whole genome data sets, the reconstruction of phylogenetic trees at a genomic scale has become feasible. Exploration of these large data sets has revealed that there may be discordance among the topologies in different genomic regions [19, 20]. Although these differences may be caused by incorrect estimations of gene genealogies, incongruent gene trees can also be attributed to the differing evolutionary histories of different genomic regions, especially for close species or subspecies. Traditionally, there are two types of phylogenetic analysis methods, the consensus method and the total evidence method. Both methods barely quantify the topological discordance across the entire genome. Recently, BCA, which is an improvement upon the consensus method, has been used to statistically quantify the discordance, as well as to generate phylogenetic trees [21]. A few studies using BCA have demonstrated its great potential for the reconstruction of phylogenic trees of mouse subspecies [11, 13, 22]. These studies indicate that BCA is a suitable method to quantify the proportions of Chr 1 sequence in B6-Chr1<sup>KM</sup> derived from the different subspecies. Through BCA, we found approximately that 90% and 10% of the sequences of Chr 1 were derived from M. m. domesticus and M. m. musculus, respectively. Although the sequence similarity analysis revealed that there were some regions which had higher sequence similarity with CAST, we did not observed the same results in the BCA. Therefore, we cannot make the



FIGURE 3: Sequence similarity between B6-Chr1<sup>KM</sup> and WSB, PWK, and CAST Chr 1. (a) Distribution of the numbers of 100 kb blocks of the B6-Chr1<sup>KM</sup> Chr 1 with sequence similarities (%) to the corresponding blocks of the WSB, PWK, and CAST Chr 1. (b) Sliding window analysis of the similarities of Chr 1 sequences between B6-Chr1<sup>KM</sup> and WSB, CAST, or PWK. The B6-Chr1<sup>KM</sup> Chr 1 sequence was compared using 500 kb windows and 100 kb sliding intervals. The horizontal line indicates the level of 99.7% sequence similarity.



FIGURE 4: Phylogenetic analysis of B6-Chr1<sup>KM</sup> Chr 1. (a) Fine-scale phylogenetic discordance of B6-Chr1<sup>KM</sup> Chr 1 (PP indicates posterior probability). Red represents WSB, green indicates PWK, and white represents unknown. (b) Phylogenetic tree of the WSB-derived or PWK-derived sequences of B6-Chr1<sup>KM</sup> and the wild-derived inbred mouse strain sequences. A neighbor-joining tree was generated using MEGA6 software. Red and green indicate regions supporting a single topology for KM/WSB and KM/PWK, respectively, which are both associated with a high posterior probability, as determined by BCA.

conclusion that some of Chr 1 sequence of B6-Chr1<sup>KM</sup> came from CAST which represent *M. m. castaneus*. While for PWK, highly conserved genomic regions (Figure 3(b)) with KM aligned well with the BCA results (Figure 4(a)). Thus, from both analyses, we can make the conclusion that Chinese KM mouse has a mosaic genome structure with sequences predominately derived from *M. m. domesticus* and with at least some of the remaining sequences derived from *M. m. musculus*.

In summary, we presented the analysis of a high-quality genome sequence of the B6-Chr1<sup>KM</sup>. These data allow better understanding of the structure and origin of the genetic variations in the B6-Chr1<sup>KM</sup> mouse strain, which provides insights into the utility of this mouse strain and the KM outbred stock for further biomedical research and the study of complex diseases.

#### **Data Access**

All raw reads were submitted to NCBI Sequence Read Archive under the Accession no. SRR2954707 associated with BioProject Accession no. PRJNA298468 and BioSample Accession no. SAMN04159475.

#### **Conflicts of Interest**

The authors declare that they have no competing interests.

#### **Authors' Contributions**

Fuyi Xu and Tianzhu Chao contributed equally to this work.

#### Acknowledgments

This work was supported by the Key Project of Science & Technology Commission of Shanghai Municipality (no. 13140900300), the National Science Foundation of China

(no. 31171199), the Fundamental Research Funds for the Central Universities (no. 2232013A3-06), and the DHU Distinguished Young Professor Program (B201308).

#### References

- X. Zhang, Z. Zhu, Z. Huang, P. Tan, and R. Z. Ma, "Microsatellite genotyping for four expected inbred mouse strains from KM mice," *Journal of Genetics and Genomics*, vol. 34, no. 3, pp. 214–222, 2007.
- [2] B. Yue, S. Liu, D. Liu et al., "Comparative studies on the genetic biological markers of five closed colonies of Kunming mice," *Laboratory Animal Science and Management*, vol. 20, no. S1, pp. 58–62, 2003.
- [3] J. Xiao, Y. Liang, K. Li et al., "A novel strategy for genetic dissection of complex traits: the population of specific chromosome substitution strains from laboratory and wild mice," *Mammalian Genome*, vol. 21, no. 7-8, pp. 370–376, 2010.
- [4] F. Xu, T. Chao, Y. Liang et al., "Genome sequencing of chromosome 1 substitution lines derived from Chinese wild mice revealed a unique resource for genetic studies of complex traits," *G3 (Bethesda)*, vol. 6, no. 11, pp. 3571–3580, 2016.
- [5] R. K. Patel and M. Jain, "NGS QC Toolkit: a toolkit for quality control of next generation sequencing data," *PloS One*, vol. 7, no. 2, article e30619, 2012.
- [6] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, 2009.
- [7] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
- [8] A. McKenna, M. Hanna, E. Banks et al., "The Genome Analysis Toolkit: a MapReduce framework for analyzing nextgeneration DNA sequencing data," *Genome Research*, vol. 20, no. 9, pp. 1297–1303, 2010.
- [9] P. Danecek, A. Auton, G. Abecasis et al., "The variant call format and VCFtools," *Bioinformatics*, vol. 27, no. 15, pp. 2156–2158, 2011.

- [10] W. McLaren, B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham, "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor," *Bioinformatics*, vol. 26, no. 16, pp. 2069–2070, 2010.
- [11] M. A. White, C. Ané, C. N. Dewey, B. R. Larget, and B. A. Payseur, "Fine-scale phylogenetic discordance across the house mouse genome," *PLoS Genetics*, vol. 5, no. 11, article e1000729, 2009.
- [12] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: molecular evolutionary genetics analysis version 6.0," *Molecular Biology and Evolution*, vol. 30, no. 12, pp. 2725–2729, 2013.
- [13] T. M. Keane, L. Goodstadt, P. Danecek et al., "Mouse genomic variation and its effect on phenotypes and gene regulation," *Nature*, vol. 477, no. 7364, pp. 289–294, 2011.
- [14] J. A. Blake, J. T. Eppig, J. A. Kadin et al., "Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse," *Nucleic Acids Research*, vol. 45, no. D1, pp. D723–D729, 2017.
- [15] J. S. Friedman, B. Chang, C. Kannabiran et al., "Premature truncation of a novel protein, RD3, exhibiting subnuclear localization is associated with retinal degeneration," *American Journal of Human Genetics*, vol. 79, no. 6, pp. 1059–1070, 2006.
- [16] J. Hammersen, J. Hou, S. Wunsche, S. Brenner, T. Winkler, and H. Schneider, "A new mouse model of junctional epidermolysis bullosa: the LAMB3 628G>A knockin mouse," *The Journal of Investigative Dermatology*, vol. 135, no. 3, pp. 921–924, 2015.
- [17] S. Shi, H. Wang, B. Cui et al., "Study on genetic variants of Chinese KM mouse subcolonies," *Chinese Journal of Laboratory Animal Science*, vol. 1, no. 1, pp. 29–36, 1991.
- [18] G. Zhao, S. Bao, D. Zhang, R. Zhang, and M. Jin, "Probing into the course of formation of genetic character of KM mouse," *Shanghai Laboratory Animal Science*, vol. 14, no. 1, pp. 1–4, 1994.
- [19] D. A. Pollard, V. N. Iyer, A. M. Moses, and M. B. Eisen, "Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting," *PLoS Genetics*, vol. 2, no. 10, article e173, 2006.
- [20] G. Giribet, G. D. Edgecombe, and W. C. Wheeler, "Arthropod phylogeny based on eight molecular loci and morphology," *Nature*, vol. 413, no. 6852, pp. 157–161, 2001.
- [21] C. Ane, B. Larget, D. A. Baum, S. D. Smith, and A. Rokas, "Bayesian estimation of concordance among gene trees," *Molecular Biology and Evolution*, vol. 24, no. 2, pp. 412–426, 2007.
- [22] T. Takada, T. Ebata, H. Noguchi et al., "The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains," *Genome Research*, vol. 23, no. 8, pp. 1329–1338, 2013.



BioMed Research International







International Journal of Genomics







Submit your manuscripts at https://www.hindawi.com





The Scientific World Journal



Archaea



Anatomy Research International









International Journal of Evolutionary Biology



Biochemistry Research International



Molecular Biology International



Advances in Bioinformatics



Journal of Marine Biology