

Limited RNA Editing in Exons of Mouse Liver and Adipose

Sandrine Lagarrigue,^{*,†,1,2} Farhad Hormozdiari,^{§,**,1} Lisa J. Martin,^{††,1} Frédéric Lecerf,^{*,†,3} Yehudit Hasin,^{††}
 Christoph Rau,^{††} Raffi Hagopian,^{††} Yu Xiao,^{††} Jun Yan,^{††} Thomas A. Drake,^{**} Anatole Ghazalpour,^{**}
 Eleazar Eskin,^{§,**,4} and Aldons J. Lusis^{**,††,§§,2}

*Agrocampus Ouest, Unité Mixte de Recherche (UMR) 1348 Physiology, Environment and Genetics for the Animal and Livestock Systems (PEGASE), Animal Genetics Laboratory, F-35000 Rennes, France, [†]Institut National de la Recherche Agronomique, UMR 1348 PEGASE, F-35000 Rennes, France, [‡]Université Européenne de Bretagne, F-35000 Rennes, France, and [§]Department of Computer Sciences, ^{**}Department of Human Genetics, ^{††}Department of Medicine/Division of Cardiology, ^{††}Department of Pathology and Laboratory Medicine, and ^{§§}Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, California 90095

ABSTRACT Several studies have investigated RNA–DNA differences (RDD), presumably due to RNA editing, with conflicting results. We report a rigorous analysis of RDD in exonic regions in mice, taking into account critical biases in RNA-Seq analysis. Using deep-sequenced F1 reciprocal inbred mice, we mapped 40 million RNA-Seq reads per liver sample and 180 million reads per adipose sample. We found 7300 apparent hepatic RDDs using a multiple-site mapping procedure, compared with 293 RDD found using a unique-site mapping procedure. After filtering for repeat sequence, splice junction proximity, unidirectional strand, and extremity read bias, 63 RDD remained. In adipose tissue unique-site mapping identified 1667 RDD, and after applying the same four filters, 188 RDDs remained. In both tissues, the filtering procedure increased the proportion of canonical (A-to-I and C-to-U) editing events. The genomic DNA of 12 RDD sites among the potential 63 hepatic RDD was tested by Sanger sequencing, three of which proved to be due to unreferenced SNPs. We validated seven liver RDD with Sequenom technology, including two noncanonical, *Gm5424* C-to-I(G) and *Pisd* I(G)-to-A RDD. Differences in diet, sex, or genetic background had very modest effects on RDD occurrence. Only a small number of apparent RDD sites overlapped between liver and adipose, indicating a high degree of tissue specificity. Our findings underscore the importance of properly filtering for bias in RNA-Seq investigations, including the necessity of confirming the DNA sequence to eliminate unreferenced SNPs. Based on our results, we conclude that RNA editing is likely limited to hundreds of events in exonic RNA in liver and adipose.

SEVERAL recent studies have investigated genome-wide RNA editing using deep sequencing of transcriptomes by RNA-Seq on human transformed cells, cancer cell lines (Ju *et al.* 2011; Li *et al.* 2011; Bahn *et al.* 2012; Peng *et al.* 2012; Ramaswami *et al.* 2012), or tissues of inbred mouse strains (Danecek *et al.* 2012; Gu *et al.* 2012). Total reported RNA–DNA differences (RDD) sites have varied from hundreds to

thousands. Over the same period, technical issues, such as mapping of reads in paralogous or repetitive sequence regions, mapping errors at splice sites, and systematic sequencing errors that could produce a large number of false-positive RDDs have been described (Kleinman and Majewski 2012; W. Lin *et al.* 2012; Pickrell *et al.* 2012). Another reported source of RDD error is undetected genomic DNA SNPs, arising from insufficient coverage of current DNA sequencing data (Schridder *et al.* 2011).

We have examined genome-wide exonic RDD by using RNA-Seq data obtained from two tissues, liver and adipose, in F1 reciprocal crosses from two inbred strains of mice, DBA/2J (D2) and C57BL/6J (B6). These inbred mouse strains have been subjected to deep genomic sequencing and SNP analyses, with a higher coverage for B6 than for D2. A major aim was to estimate the impact of the major technical

Copyright © 2013 by the Genetics Society of America

doi: 10.1534/genetics.112.149054

Manuscript received December 23, 2012; accepted for publication January 28, 2013

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.149054/-/DC1>.

¹These authors contributed equally to this work.

²Corresponding authors: Agrocampus Ouest Department of Animal Sciences, 65 rue de Saint Briec, Rennes, 35000 France. E-mail: Sandrine.lagarrigue@agrocampus-ouest.fr; and UCLA Department of Medicine/Division of Cardiology, 3730 MRL, 650 Charles E. Young Dr. South, Los Angeles, CA 90095-1679. E-mail: jlusis@mednet.ucla.edu

issues (paralog mapping, mismapping near splice sites and repeat sequences, and systematic sequencing errors, such as unidirectional strand and extremity biases) to obtain a better sense of the true frequency of RDD in normal mammalian tissues. The RDDs that remained were then characterized by comparison with expressed sequence tags and tested by Sanger and quantitative Sequenom sequencing, showing the importance of controlling the genomic DNA sequence in RDD site analysis. We also examined the effects of sex and diet and the possibility of allele-specific RNA editing.

Materials and Methods

Ethics statement

All animals were handled in strict accordance with good animal practice as defined by the relevant national and/or local animal welfare bodies, and all animal work was approved by the appropriate committee. All experiments in this article were carried out with UCLA IACUC approval.

Mice and tissues

RNA-Seq was performed on liver and adipose mRNA from F1 male and female D2 and B6 mice, purchased from the Jackson Laboratory (Bar Harbor, ME). Reciprocal F1 male and female mice were generated by breeding the parental strains in the vivarium at University of California, Los Angeles (UCLA). For six liver RNA libraries, RNA from three mice was pooled into four independent samples of high-fat-fed B6xD2 (BXD) and DXB males and females and two samples of chow fed BXD and DXB males. Four adipose RNA libraries were made using pooled RNA from three BXD and DXB males and females fed a chow diet. Males and females of other reciprocal inbred mouse crosses were used for Sequenom validation. Those F1's were A/JxC3H/HeJ (AXH) and HXA and B6xC3H/HeJ (BxH) and HXB. Liver RNA was isolated from three mice per sex per F1 cross using the RNeasy kit from Qiagen (Valencia, CA). cDNA was made with the High-Capacity Reverse Transcription kit from Applied Biosystems. All mice were fed *ad libitum* and maintained on a 12-hr light/dark cycle. F1 pups were weaned at 28 days and fed a chow diet (Ralston-Purina Co.) until 8 weeks of age, at which time half were placed on a high-fat diet (Research Diets D12266B). All F1 mice were killed at 16 weeks, with liver and adipose harvested at that time.

Library preparation for Illumina sequencing

Library preparation was performed as recommended by the manufacturer (Illumina, Hayward, CA). Briefly, total RNA was extracted using the RNeasy Mini kit with DNase treatment (Qiagen, Valencia, CA). Poly(A) mRNA was isolated and fragmented, and first-strand cDNA was prepared using random hexamers. Following second-strand cDNA synthesis, end repair, addition of a single A base, adaptor ligation, and agarose gel isolation of ~200-bp cDNA, PCR amplification of the ~200-bp cDNA was performed. Liver samples were sequenced using the Illumina GAII sequencer to a coverage of ~40 million single-end 75-bp reads per sample. Adipose was sequenced

with the Illumina HiSeq2000 on paired-end 50-bp reads and generated 180 million reads per sample.

Read mapping

We first aligned reads 75 bp (liver) or 50 bp (adipose) to the mouse reference genome version mm9 using mrsFAST (Hach *et al.* 2010), allowing up to five mismatches for liver tissue and three for adipose. The reads were divided into two categories. The first category was the set of mapped reads that align to the genome with e or less mismatches. The second category was reads that failed to map to the reference genome. Many RNA-Seq reads failed to align to a genome because they spanned the exonic junctions. To overcome this problem we mapped the unmapped reads with TopHat (Trapnell *et al.* 2009), which is designed to map reads to the genome by splitting the reads into smaller fragments. The reads aligned to the genome in this process were added to the map read set. In our multiple-site mapping procedure, we allowed up to 10 genomic locations for the same read. If a read mapped to more than 10 locations, we picked the top 10 sites with the most reads. In the unique-site mapping procedure, we consider only the reads that mapped to only one position, and reads that map to more than one position are filtered out.

RDD criteria

We selected reads with base modifications of the RNA represented by at least 10 reads at the position for the edited base, located in one exon, and not corresponding to a known genomic SNP between D2 and B6 (based on the Mouse Sequencing Consortium; Waterston *et al.* 2002; Keane *et al.* 2011). The read was required to have a base quality of 20 or higher. A base modification A (DNA base) → B (edited base) was considered an RDD if (1) B6 and D2 were homozygous for the DNA nucleotide in the genomic DNA, (2) B generated at least 10 RNA-Seq reads, with (3) $(A + B)/\text{total reads} \geq 0.9$ of all reads for the site and $B/(A + B) \geq 0.1$ in RNA-Seq, (4) the RDD was shared by four of four samples for adipose tissue or four of six samples for liver (supporting information, Figure S1). Filter 2 ensured a certain level of expression for the transcript with the edited base. The last filter allowed exclusion of random sequencing errors (Meacham *et al.* 2011). The simple sequence repeats (SSR) patterns were investigated within the sequences using SciRoKo (Kofler *et al.* 2007) with the following parameters: perfect repeats mode, minimum repeats of 3, and a minimum length of pattern of 5 bases. The SSR patterns were investigated near the RDD site with an offset of ± 3 bases. To eliminate unidirectional strand bias, an RDD was kept only if the proportion of reads belonging to the forward sequencing strand was $\geq 20\%$ and $\leq 80\%$, in at least three of four samples for adipose and four of six samples for liver. To eliminate the sequencing extremity bias, an RDD was kept only if the proportion of reads where the alternative base was found in the five first or last 5 bases of the sequencing read was $\leq 50\%$ of the reads in at least three samples among four for adipose and four samples among six for liver. The thresholds for the

unidirectional strand and sequencing extremity bias have been calculated according to the intersection of the distributions of RDD and SNP sites (see Figures 2 and 3).

EST analysis

cDNA sequences with 100 bp flanking the edited base were extracted from the University of California, Santa Cruz (UCSC) database. Using these 201-bp sequences containing the edited base in the middle, BLAST analyses were performed against the mouse EST database. For each RDD site we counted the number of ESTs with the DNA base and with the edited base. All BLAST analyses were conducted with the default parameters except the gap open and extend costs fixed to 200 (to avoid gaps). The BLAST results were filtered by lengths of the alignments ≥ 75 bp and five or more mismatches.

Sex and diet effects

An RDD site was declared as affected by sex or diet if the fold change “Edited base reads/DNA base reads” between the two diets or the two sexes is >1.5 (or $<1/1.5$). The significance was tested by a Fisher exact test using *P*-values corrected for multiple testing by the Benjamini–Hochberg method. A cutoff of 0.05 was then used.

Sequenom validation

DNA and RNA were extracted from three mice per cross (independent sample from RNA-Seq RNA), pooled, and cDNA generated from the RNA pool. DNA and cDNA were analyzed in a primer extension assay, designed to target the RDD sequence. The primer extension assay was carried out using the MassARRAY (Sequenom iPLEX Gold genotyping protocol) platform according to the manufacturer’s specifications by the McGill University and Génome Québec Innovation Centre. Primer extension products were analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF/MS). If the RDD were present, it would generate two peaks on the MS profiles. The area of each peak was proportional to the transcript abundance and was measured by the MassARRAY software to generate an “edited base/reference base” ratio calculation. The SNP at the cDNA level was compared to the genomic DNA at the same position, the latter being expected to be homozygous for a true RDD site.

Sanger sequencing

PCR primers for genomic DNA, designed using the Primer3 website from the Massachusetts Institute of Technology (MIT), produced a PCR product that was at least 400 bases upstream and downstream of the RDD on the DNA. Nested primers were designed and used to produce a product that was at least 200 bases on each side of the RDD site. The nested PCR procedure was necessary due to multiple PCR product bands, sometimes overlapping, before this was implemented. The PCR product was run on an agarose gel, cut out, and purified by QIAquick Gel Extraction kit by Qiagen. Sanger sequencing was performed by the UCLA Sequencing Core.

Results

RNA editing

We investigated the frequency of RNA editing in primary mouse tissues by mapping 232 million RNA-Seq generated reads from liver and 723 million reads from adipose of F1 mice of inbred strains B6 and D2. (See Figure S1 for flowchart.) We analyzed six independent liver and four independent adipose samples from BXD and DXB males and females, with each sample containing pooled RNA from three mice (Table S1).

From hepatic RNA-Seq reads, we initially identified 7319 exonic RDD sites with a multiple mapping procedure, which allowed for the mapping of a read to more than one location in the genome. We analyzed the same data using a unique mapping procedure and reduced the number to 293 exonic RDDs, indicating the strong impact of paralog alignments. We then filtered for mismapping bias near a splice junction (7%), sequence repeats proximity (27%) (Figure 1), unidirectional bias (49%), and sequencing extremity bias (56%) (Figure 1, Figure 2, and Figure 3). After applying these four filters, 63 RDDs remained for liver (Figure 1 and Table S2). For adipose tissue, the filter for unique mapping was applied from the outset, resulting in 1667 exonic adipose RDDs, and this number was reduced to 188 after applying the filters for mapping and sequencing bias (Figures 1–3 and Table S3). In both liver and adipose, unidirectional strand and sequencing extremity biases were the major factors generating false-positive RDDs. The difference of unidirectional bias between the tissues, 49% liver and 73% adipose, and sequencing extremity bias, 56% liver and 38% adipose, is possibly due to the sequencing method, Illumina GAIIIX on 75 bp reads for liver and HiSeq2000 on 50 bp reads for adipose. The 63 RDDs in liver represented 32 genes, and 188 RDDs in adipose represented 86 genes. The higher number of RDDs found in adipose compared to liver probably reflects, at least in part, the higher ($\sim 5\times$) number of reads generated for adipose. Fifteen exonic RDD sites representing 9 genes were shared (Table S2 and Table S3), suggesting that RNA editing is largely tissue specific.

RDD analysis using expressed sequence tags (ESTs)

Using an *in silico* approach, we analyzed the exonic RDD sites with the public mouse EST database (v. 128). Liver (58%) and adipose (55%) RDDs appeared at least once in the EST database, and of those, 46% in liver and 42% in adipose, the edited base ESTs/total ESTs ratio was ≥ 0.10 (Figure 1). Therefore, approximately half of the RDDs that we found have been produced by other sequencing and RNA isolation methods, giving credence to those sites as real RDD. On the other hand, half of the potential RDDs identified in our present study are novel. Our data indicate that this could be due either to low expression of the mRNA containing the edited base, resulting in its absence from the EST database, or to unaccounted-for error in mapping or sequencing.

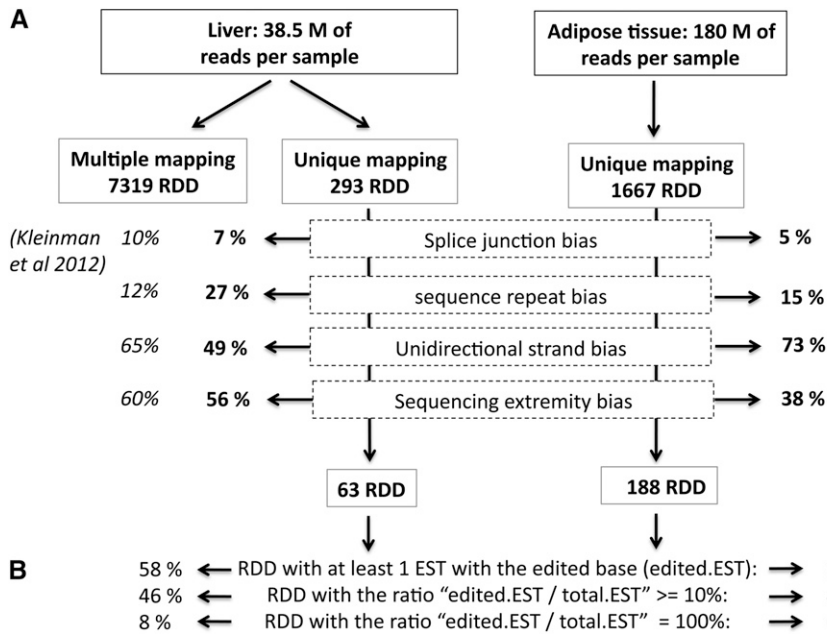


Figure 1 RDD in mouse liver and adipose identified by RNA-Seq. (A) RDD numbers resulting from different filters. The results found for the same filters by Kleinman and Majewski (2012) are in italics. (B) Percentages of RDD reads found in comparison with reads from the EST database.

RDD characteristics

We characterized the resulting RDD according to RNA editing category (Figure 4). In the original 293 liver and 1667 adipose RDD obtained by the unique mapping procedure, we found more transversions (purine to pyrimidine or pyrimidine to purine), 60% for liver and 65% for adipose, than transitions (purine to purine or pyrimidine to pyrimidine) (Figure 4). However, in the final 63 RDD for liver and 188 for adipose, transitions were more abundant in both tissues, 62 and 66%, respectively (Figure 4). A-to-I(G) and C-to-U(T) canonical editing events (*i.e.*, deaminase enzyme dependent) represented 56 and 52% of the transitions in liver and adipose, respectively. The noncanonical categories, observed at a lower frequency, may represent novel editing mechanisms, opposite-strand transcripts, or other unknown sources of error (see *Discussion*).

RDD validation

Due to genetic drift or to low sequencing coverage, some SNPs that may not be referenced in SNP databases exist. Because these SNPs would lead to spurious RDD sites, we tested by Sanger sequencing the genomic DNA from 12 RDD among the potential 63 hepatic RDD. We found that three of those RDD sites resulted from an unreferenced SNP. We also tested 14 RDD with Sequenom technology (Table 1 and Table S2). We made new cDNA from different sets of mice to avoid duplication of reverse transcription errors. Seven RDD sites that remained after filtering, corresponding to four genes, were selected to represent a range of edited base ratios (EBR) from 0.33 to 1 in favor of the edited base (Table 1). An RDD site was confirmed by Sequenom if no polymorphism was observed in the genomic DNA in the F1 mice, but was observed in the cDNA. All seven potentially real RDD sites in liver were confirmed (Table 1, Figure 5, and Table S4). In

those, we found similar proportions of reads with the edited base in Sequenom and RNA-Seq technologies (Figure 5, A and B). The other seven RDD sites, chosen from those that

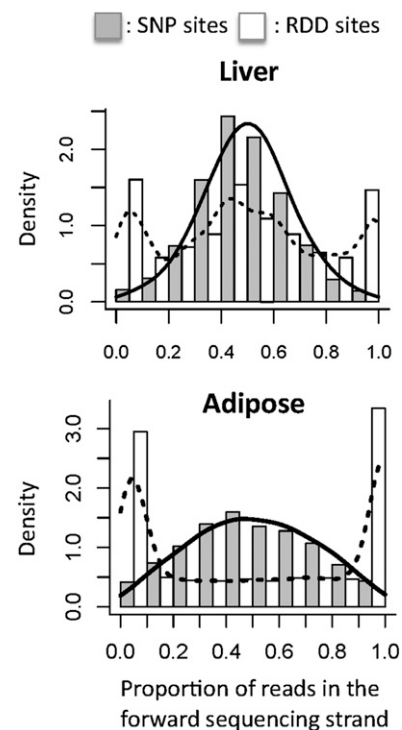


Figure 2 Contribution of sequencing strand bias to RDD. Histogram with the distribution of RDD (y-axis) according to the proportion of reads belonging to the forward sequencing strand (x-axis) in liver (sample M.CH.BxD) and adipose (sample F.BxD). Open bars and the dashed line correspond to the RDD sites. Shaded bars and the solid line correspond to the dbSNP sites that are polymorphic in the sample and used as a control data set. The false positives are calculated using the tails of the distributions where the RDD and SNP distributions intersect. The same shape of distribution was observed for all samples of the same tissue.

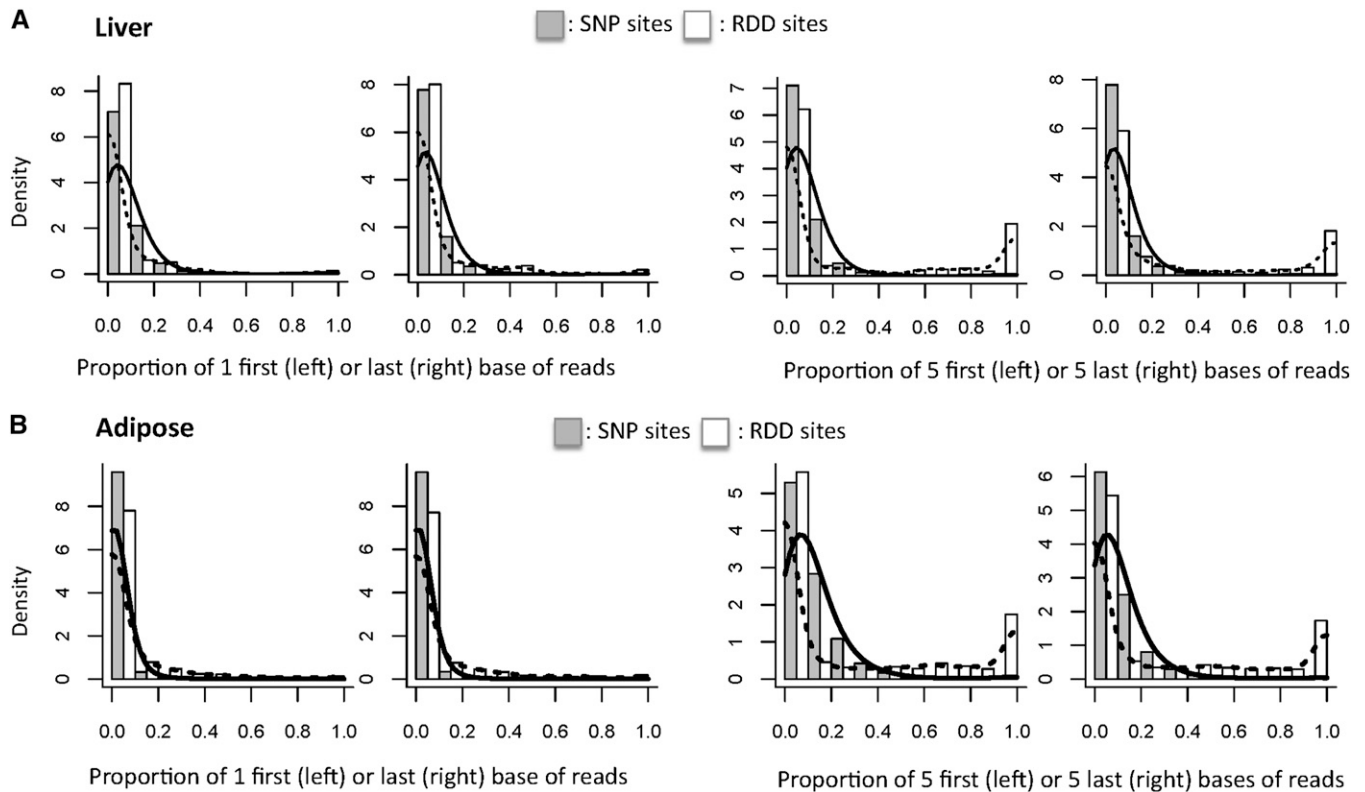


Figure 3 Contribution of end of read sequencing to RDD. Histogram with the distribution of RDD (y-axis) according to the proportion of reads where the alternative base is found in the first or last base or in the 5 first or 5 last bases of the sequencing read (x-axis). (A) Liver (sample M.CH.BxD) and (B) adipose (sample F.BxD). Open bars and the dashed line correspond to the RDD sites. Shaded bars and the solid line correspond to the dbSNP sites that are polymorphic in the sample and used as a control data set. The false positives are calculated using the tails of the distributions where the RDD and SNP distributions intersect. The same shape of distribution was observed for all samples of the same tissue.

had been filtered out, were confirmed as true negatives by Sequenom analysis (Table 1). The degree of editing in the seven RDD that we validated by Sequenom ranged from nearly 100% in *Gm5424* to ~20% in *Pisd* (Figure 5, Table S4, and Figure S2). *Npm3-ps1* and *Gm5424* are categorized as pseudogenes without protein products. Notably, two of the RDD that were validated by Sequenom, *Gm5424* C-to-G and *Pisd* G-to-A, were noncanonical, and to be certain of the results, we also confirmed these RDD by Sanger sequencing (Figure S3).

Genetic regulation of RNA editing and analysis of diet and sex effects

Evidence of genetic control was detected in the Sequenom validation studies in which F1 reciprocal crosses of AXH and BXH, were analyzed along with the BXD F1 reciprocal cross mice. We found similar EBR for all the RDD among the different genetic backgrounds, except for *Npm3-ps1* (Figure 5). The expression pattern for *Npm3-ps1* (Figure 5A) suggested that RDD can be influenced by genetics, sex, and diet. First, we observed a lower EBR of this gene in AXH and HXA F1 mice compared to the other crosses that were not due to a DNA SNP in this strain (Figure 5A and Table S4). Second, we observed a significantly higher EBR of AXH and HXA males fed a high-fat diet compared with the EBR of females

fed a high-fat diet or males fed a chow diet on the same genetic background. However, we did not observe any RDD sites affected by diet and sex in the 63 RNA-Seq hepatic

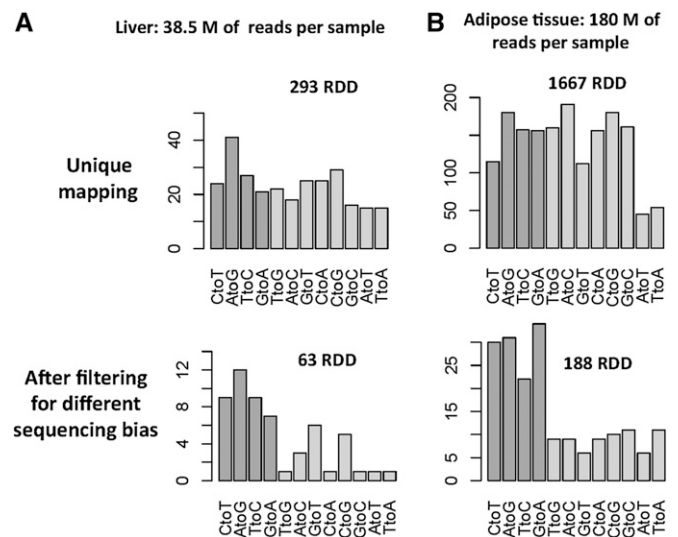


Figure 4 RDD categories in liver and adipose tissue. (A) RDD categories observed with liver RNA-Seq data after the multiple and unique mapping procedures and filtering for sequencing bias. (B) RDD categories observed with adipose RNA-Seq data after the unique mapping procedure and filtering for sequencing bias.

Table 1 RDD tested by Sequenom mass spectrometry

Gene	RDD on coding strand	Chromosome	Genomic location	Edited base ratio	Yes if remaining after applying bias corrections; specific bias if not	Validated by Sequenom
Gm5424	C-to-T	10	61534269	1	Yes	Yes
Gm5424	C-to-T	10	61534299	1	Yes	Yes
Gm5424	C-to-G*	10	61534526	0.99	Yes	Yes
Pisd	A-to-G	5	33079182	0.61	Yes	Yes
Pisd	G-to-A*	5	33079194	0.63	Yes	Yes
Npm3-ps1	C-to-T	6	85026710	0.54	Yes	Yes
Serinc1	C-to-T	10	57235791	0.33	Yes	Yes
Knng1	G-to-A*	16	23079125	0.19	read.extremity_bias	No
Fxyd1	T-to-G*	7	31839392	0.28	read.extremity_bias	No
Fxyd1	A-to-G	7	31839391	0.18	read.extremity_bias	No
ApoA2	A-to-G	1	173155222	0.15	read.extremity_bias	No
Apoc1	C-to-A*	7	20277841	0.74	Unidirectional strand bias	No
Ttr	C-to-A*	18	20832387	0.51	Unidirectional strand bias	No
2610027L16Rik	C-to-T	14	56373961	0.87	Unidirectional strand bias	No

Seven RDD that remained after filtering were validated. Seven that had been filtered out were confirmed as true negatives. U is represented by T and I is represented by G. Noncanonical editing is designated by an asterisk (*).

RDD, using the BXD and DXB reciprocal crosses as biological replicates for each sex and diet (Table S1), suggesting that effects of these factors on RDD are modest.

Discussion

We have carried out whole-genome analysis of RDD in liver and adipose in mice to examine the extent and nature of the

phenomenon and to explore potential effects of sex, diet, and genetics. The inbred mouse strains, B6 and D2, have been deep sequenced and studied for SNPs, helping to avoid issues such as DNA heterogeneity that likely confound studies with human cell lines. Using multiple-site mapping at genomic sites, we obtained 7319 apparent RDD sites in mouse liver, similar in frequency to the initial report of

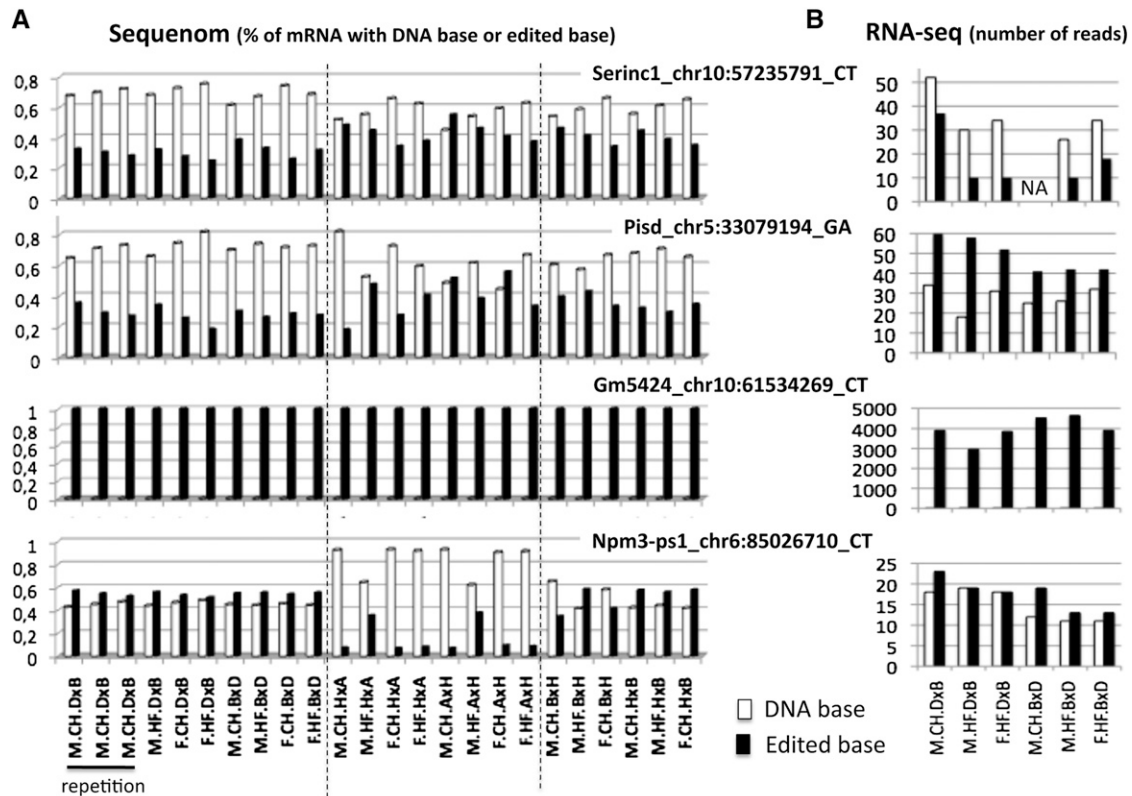


Figure 5 Validation by Sequenom technology of four RDD observed by RNA-Seq in liver. (A) Results obtained by Sequenom technology and expressed as percentage of total mRNA sequences containing the DNA base (open) vs. the edited base (solid) (total, 100%). (B) Results obtained by RNA-Seq technology and expressed as number of reads mapping to the RDD position.

extensive RDD in human B cells (Li *et al.* 2011). However, after rigorous filtering for mapping and sequencing bias errors, our RDD total was 63 in liver and 188 in adipose. Our comprehensive analysis is relevant to the interpretation of RNA-Seq results that detect polymorphisms, for example, in allele-specific expression studies.

A major aim of our study was to determine the importance of five known RNA-Seq biases on RDD investigations. The most significant bias in our study was mismapping to paralogous locations, the elimination of which reduced the RDD total from thousands to hundreds. Following unique mapping, the other important biases were unidirectional strand and extremity read bias, which filtered out 40–73% false-positive RDD. Filters for splice junction and sequence repeats bias eliminated between 5 and 27% of RDD. Our results were consistent with those previously reported (Kleinman and Majewski 2012) (Figure 1). Furthermore, we report the importance of unreferenced SNPs that could arise from low sequence coverage or spontaneous mutations, which have become fixed in a breeding colony. If we consider a mutation rate of about 2×10^{-8} /generation (Sun *et al.* 2012), and that the mice we sequenced were many generations removed from the reference-sequenced mice, a significant number of polymorphisms would be expected. Therefore, to make a legitimate claim of widespread RNA editing, it is important to apply filters for mapping and sequencing bias and to confirm the genomic DNA sequence at each RDD position. We conclude that among the 63 and 188 RDD sites found in liver and adipose, a significant portion are possibly false positives due to unreferenced genomic SNPs.

For the RDD that we identified, ~65% were transitions, and of those, ~55% were canonical A-to-I(G) and C-to-U(T) edits. Therefore, while we found a significant number of canonical RDD, likely produced by known RNA editing deaminase enzymes, we also found many novel RDD that were noncanonical. All categories of RDD have been found by other researchers (Ju *et al.* 2011; Bahn *et al.* 2012), and although no enzymes that can mediate these events have been identified, noncanonical RDD sites have been validated previously (Bahn *et al.* 2012). Some noncanonical G-to-A and T-to-C transitions may exist because regions of unknown sense–antisense transcription may lead to confusion about which is the biologically relevant read (Bahn *et al.* 2012). We investigated 14 RDD by Sequenom technology, including seven that remained after filtering and seven that were eliminated by filtering. The seven RDD that had been filtered out, three canonical and four noncanonical (T-to-G, G-to-A, and two C-to-A), were confirmed as true negatives. Of the seven RDD that were confirmed as true positives, five were canonical, one was a noncanonical C-to-G transversion (*Gm5424*), and one was a noncanonical G-to-A transition (*Pisd*). Sequenom and Sanger sequencing, which investigated differences in both genomic DNA and cDNA sequences, provided convincing evidence that these noncanonical RDD are genuine. The above possible explanation, regarding unknown sense–antisense transcription, does not apply to the *Pisd* G-to-A RDD, because the canonical A-to-G RDD on the same read would be a noncanonical T-to-C in the

reverse direction. Therefore, either unknown mechanisms for RNA editing exist, or there are unknown errors in the sequence interpretations.

The disparity of our numbers with human studies that report much larger numbers could be explained two ways. First, heterogeneity in the DNA is much less controlled for in humans. Since the human reference sequence used for mapping does not reflect the real diversity in the genome of each individual, an overestimation of RDD is likely. Second, in human studies that used transformed cell lines (Li *et al.* 2011; Bahn *et al.* 2012; Peng *et al.* 2012), high RNA editing findings may reflect a real biological change. Deaminase enzymes, that are known to be involved in editing (Conticello 2008; Nishikura 2010; Wulff and Nishikura 2010), may become more active in the transformation process, and increased RNA editing has been found in some cancer cell lines (Galeano *et al.* 2012). Equally important, viruses that were used to transform human lymphocytes may have left remnants of viral DNA in those cells (Z. Lin *et al.* 2012), and as increased “errors” in RNA and DNA are an important mechanism for virus survival (Domingo 2011; Z. Lin *et al.* 2012), the very large numbers for RDD in those cells may reflect a biological artifact.

Nevertheless, the RDD numbers reported in our study are also lower compared to two recent studies of exonic RNA editing in primary mouse tissues (Danecek *et al.* 2012; Gu *et al.* 2012). Danecek *et al.* (2012) reported ~700 exonic editing sites in whole brain of 15 mouse strains. ADAR expression and activity have been shown to be higher in brain than in other tissues, which could explain some of the discrepancy (Paul and Bass 1998); nonetheless, sequence repeat bias, which accounted for 27% of the error in our liver RDD, were apparently not filtered out in this study. Random sequencing errors were also more stringently dealt with in our study, in which the RDD occurrence in four of six samples was required, compared with that study’s requirement of two biological replicates. These authors showed good reproducibility of EBR between strains; however, unreferenced SNPs could be shared by these genetically related strains. This study used Sequenom to test 611 sites, but used genomic DNA from only one strain, C57BL/6, to confirm DNA homozygosity for all strains. Our Sanger sequencing results, in which 3 out of 12 RDD resulted from unreferenced SNPs, contradict the assumption that errors in the reference (C57BL/6) genomic sequence are not a significant source of bias.

Another recent RNA editing study in mice by Gu *et al.* (2012) reported 140 RDD sites in liver, closer in magnitude to the RDD that we found, although still higher. One difference was a lower requirement for a site to be designated as edited: whereas we required at least 10 edited base reads, and an EBR ≥ 0.1 in four of six samples, the Gu *et al.* (2012) study required two edited base reads with an EBR ≥ 0.05 in each of three biological replicates. Thirty-eight of the reported 140 RDD would have passed our more stringent requirements. Surprisingly only three of those were common

to our data set, *Slc7a2*, *Serinc1*, and *2810407C02Rik*, highlighting the difficulty of identifying a reproducible tissue-specific RDD list. Lack of agreement among studies is due in part to differences in the multistep analysis procedure, including the number of mismatched bases authorized for the read alignments, the criteria to define an RDD site, and the filters used for the different sequencing biases. It should be noted that RNA-Seq, used in our study, detected RDD in exonic regions, but did not allow for analysis of intronic or noncoding regions, and thus edited sites in these regions, which have been shown to influence splicing (Lev-Maor *et al.* 2008), would not have been detected.

We found 25% hepatic RDD sites that were common to adipose. These results are comparable to those found by Danecek *et al.* (2012), who found on average 50–60% of the RDD observed in brain common with six other tissues, demonstrating a significant level of tissue specificity of the edited sites.

Finally, our results were similar across all F1 heterozygous strains for six of the seven RDD sites that were interrogated by Sequenom, with substantial reproducibility of EBR between strains. This finding was similar to Danecek *et al.* (2012) who reported RDD shared among nine or more strains, with a good reproducibility of EBR between strains. One of the RDD confirmed in our study, *Npm3-ps1*, had a different EBR among the F1 strains, indicating an effect of genetics. In addition, within HXA and AXH samples, males fed a high-fat diet had a higher EBR for the *Npm3-ps1* RDD, suggesting an effect of sex and diet. However, an analysis of the effect of sex and diet on 63 hepatic RDD sites using BXD and DXB RNA-Seq data shows a very modest impact of these two factors.

In conclusion, our study represents a rigorous analysis of possible sources of error in whole-genome evaluations of RNA editing in mice using RNA-Seq technology in normal tissues. Our findings underscore the importance of properly filtering RNA-Seq data, not just for RNA editing investigations, but also for all applications that require the identification of polymorphisms in large data sets, such as allele-specific expression. Furthermore, we emphasize the necessity of systematically controlling the genomic homozygous status. In contrast to several recent studies using transformed human cell lines that found thousands of RNA editing events, we directly analyzed inbred mouse tissues, thereby avoiding possible error caused by genetic heterogeneity in humans. We conclude that exonic RNA editing in mouse liver and adipose is limited to hundreds or fewer RDD sites. We do find evidence for noncanonical RNA editing, in agreement with previous studies. This clearly requires further investigations, although at this point, we cannot rule out the possibility of unknown sources of sequencing error. We also find that sex and diet have relatively modest effects on RNA editing.

Acknowledgments

Funding was provided by the National Institutes of Health (NIH) grants HL28481 and HL30568 to A.J.L., and DK072206 to A.J.L. and T.A.D. F.H. and E.E. are supported by National

Science Foundation grants 0513612, 0731455, 0729049, 0916676, and 1065276 and NIH grants HL080079 and DA024417. S.L. and F.L. were supported by grants from the French genomic agricultural society (AGENAVI), INRA, and the Agence Nationale de la Recherche (grant no. 0426). HD07228 provided funding for L.J.M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

- Bahn, J. H., J. H. Lee, G. Li, C. Greer, G. Peng *et al.*, 2012 Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* 22: 142–150.
- Conticello, S. G., 2008 The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 9: 229.
- Danecek, P., C. Nellaker, R. E. McIntyre, J. E. Buendia-Buendia, S. Bumpstead *et al.*, 2012 High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.* 13: r26.
- Domingo, E., 2011 Paradoxical interplay of viral and cellular functions. *Viruses* 3: 272–277.
- Galeano, F., S. Tomaselli, F. Locatelli, and A. Gallo, 2012 A-to-I RNA editing: the “ADAR” side of human cancer. *Semin. Cell Dev. Biol.* 23(3): 244–250.
- Gu, T., F. W. Buaas, A. K. Simons, C. L. Ackert-Bicknell, R. E. Braun *et al.*, 2012 Canonical A-to-I and C-to-U RNA editing is enriched at 3’UTRs and microRNA target sites in multiple mouse tissues. *PLoS ONE* 7: e33720.
- Hach, F., F. Hormozdiari, C. Alkan, I. Birol, E. E. Eichler *et al.*, 2010 mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat. Methods* 7: 576–577.
- Ju, Y. S., J. I. Kim, S. Kim, D. Hong, H. Park *et al.*, 2011 Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. *Nat. Genet.* 43: 745–752.
- Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong *et al.*, 2011 Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477: 289–294.
- Kleinman, C. L., and J. Majewski, 2012 Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science* 335: 1302.
- Kofler, R., C. Schlotterer, and T. Lelley, 2007 SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23: 1683–1685.
- Lev-Maor, G., O. Ram, E. Kim, N. Sela, A. Goren *et al.*, 2008 Intronic Alus influence alternative splicing. *PLoS Genet.* 4: e1000204.
- Li, M., I. X. Wang, Y. Li, A. Bruzel, A. L. Richards *et al.*, 2011 Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333: 53–58.
- Lin, W., R. Piskol, M. H. Tan, and J. B. Li, 2012 Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science* 335: 1302e.
- Lin, Z., A. Puetter, J. Coco, G. Xu, M. J. Strong *et al.*, 2012 Detection of murine leukemia virus in the Epstein-Barr virus-positive human B-cell line JY, using a computational RNA-Seq-based exogenous agent detection pipeline, PARSES. *J. Virol.* 86: 2970–2977.
- Meacham, F., D. Boffelli, J. Dhahbi, D. I. Martin, M. Singer *et al.*, 2011 Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics* 12: 451.
- Nishikura, K., 2010 Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79: 321–349.
- Paul, M. S., and B. L. Bass, 1998 Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* 17: 1120–1127.

- Peng, Z., Y. Cheng, B. C. Tan, L. Kang, Z. Tian *et al.*, 2012 Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* 30: 253–260.
- Pickrell, J. K., Y. Gilad, and J. K. Pritchard, 2012 Comment on “Widespread RNA and DNA sequence differences in the human transcriptome.” *Science* 335: 1302.
- Ramaswami, G., W. Lin, R. Piskol, M. H. Tan, C. Davis *et al.*, 2012 Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* 9: 579–581.
- Schrider, D. R., J. F. Gout, and M. W. Hahn, 2011 Very few RNA and DNA sequence differences in the human transcriptome. *PLoS ONE* 6: e25842.
- Sun, J. X., A. Helgason, G. Masson, S. S. Ebenesersdottir, H. Li *et al.*, 2012 A direct characterization of human mutation based on microsatellites. *Nat. Genet.* 44: 1161–1165.
- Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.
- Waterston, R. H., K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril *et al.*, 2002 Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562.
- Wulff, B. E., and K. Nishikura, 2010 Substitutional A-to-I RNA editing. *Wiley Interdiscip. Rev. RNA* 1: 90–101.

Communicating editor: D. A. Largaespada

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.149054/-/DC1>

Limited RNA Editing in Exons of Mouse Liver and Adipose

**Sandrine Lagarrigue, Farhad Hormozdiari, Lisa J. Martin, Frédéric Lecerf, Yehudit Hasin,
Christoph Rau, Raffi Hagopian, Yu Xiao, Jun Yan, Thomas A. Drake, Anatole Ghazalpour,
Eleazar Eskin, and Aldons J. Lusis**

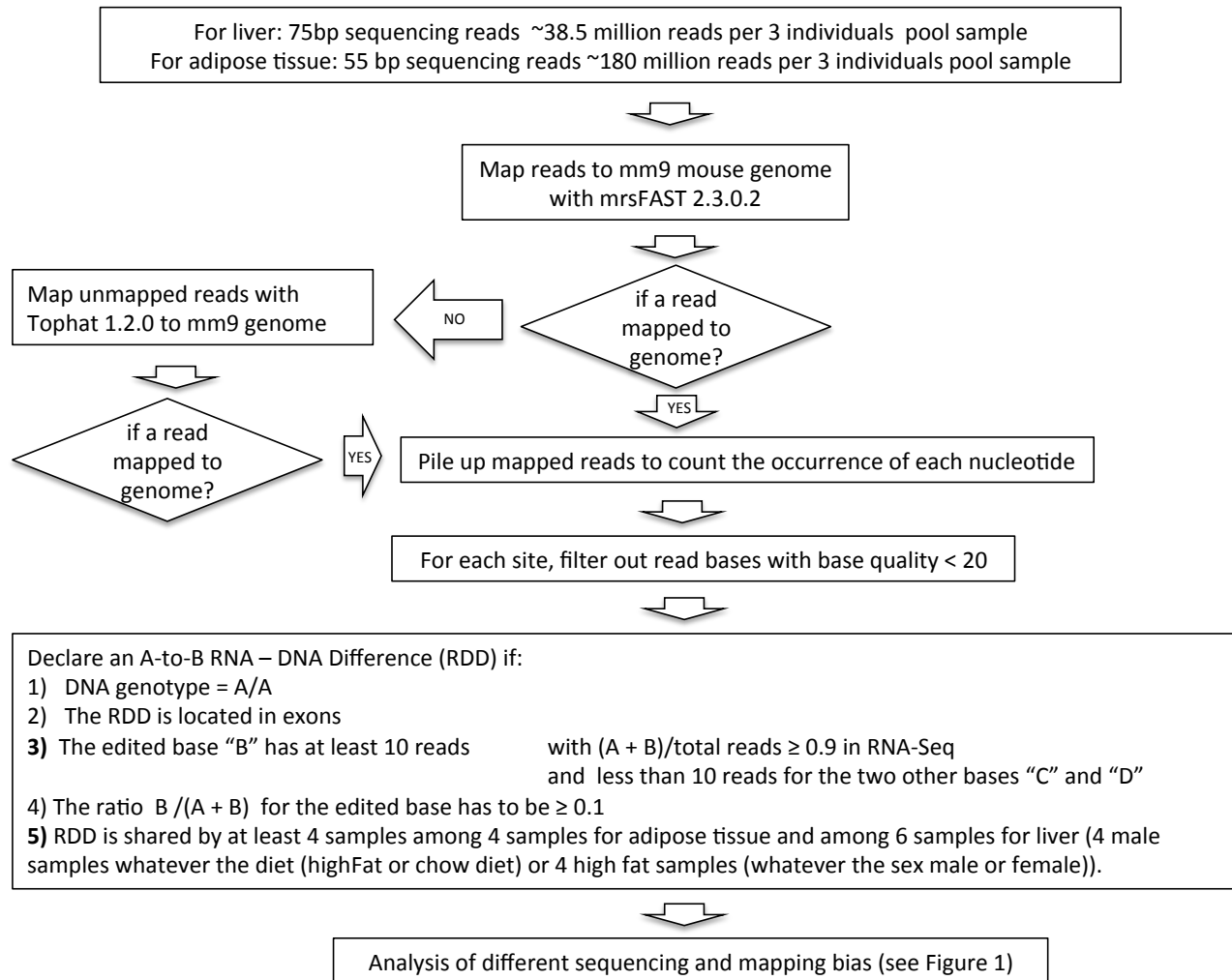


Figure S1 Flowchart of analysis

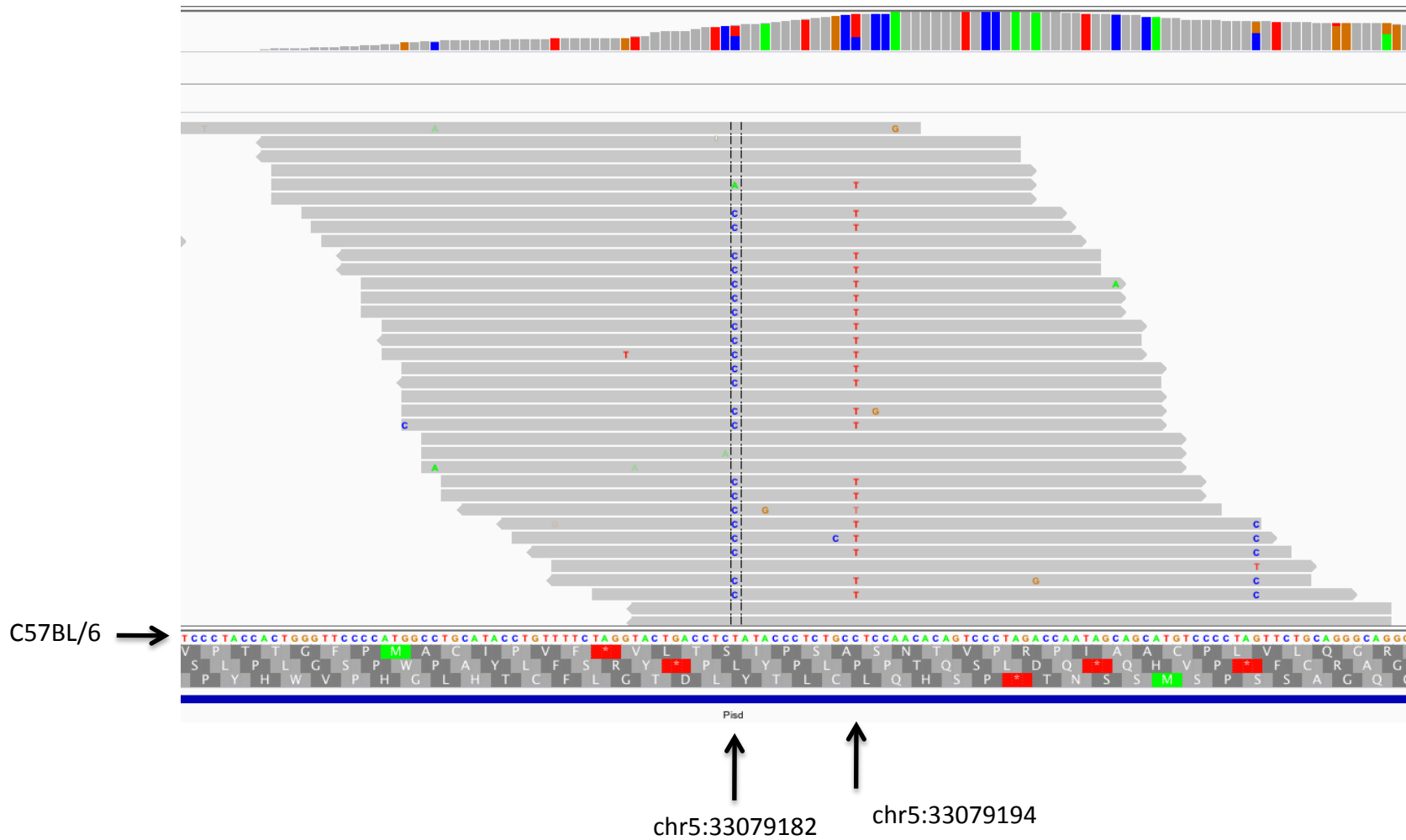
RNA-Seq read alignment
for:

Gene	Edited base on strand +	Edited base on gene strand	Chr.	Genomic location	Edited base ratio
Gm5424	CG	CG	10	61534526	0.99
Pisd	TC	AG	5	33079182	0.61

Figure S2A Here we have presented the read alignments for Gm5424 chr 10 61534526 and Pisd chr5 33079182, showing the position of the RDD within the reads and the numbers of edited base versus DNA base reads.

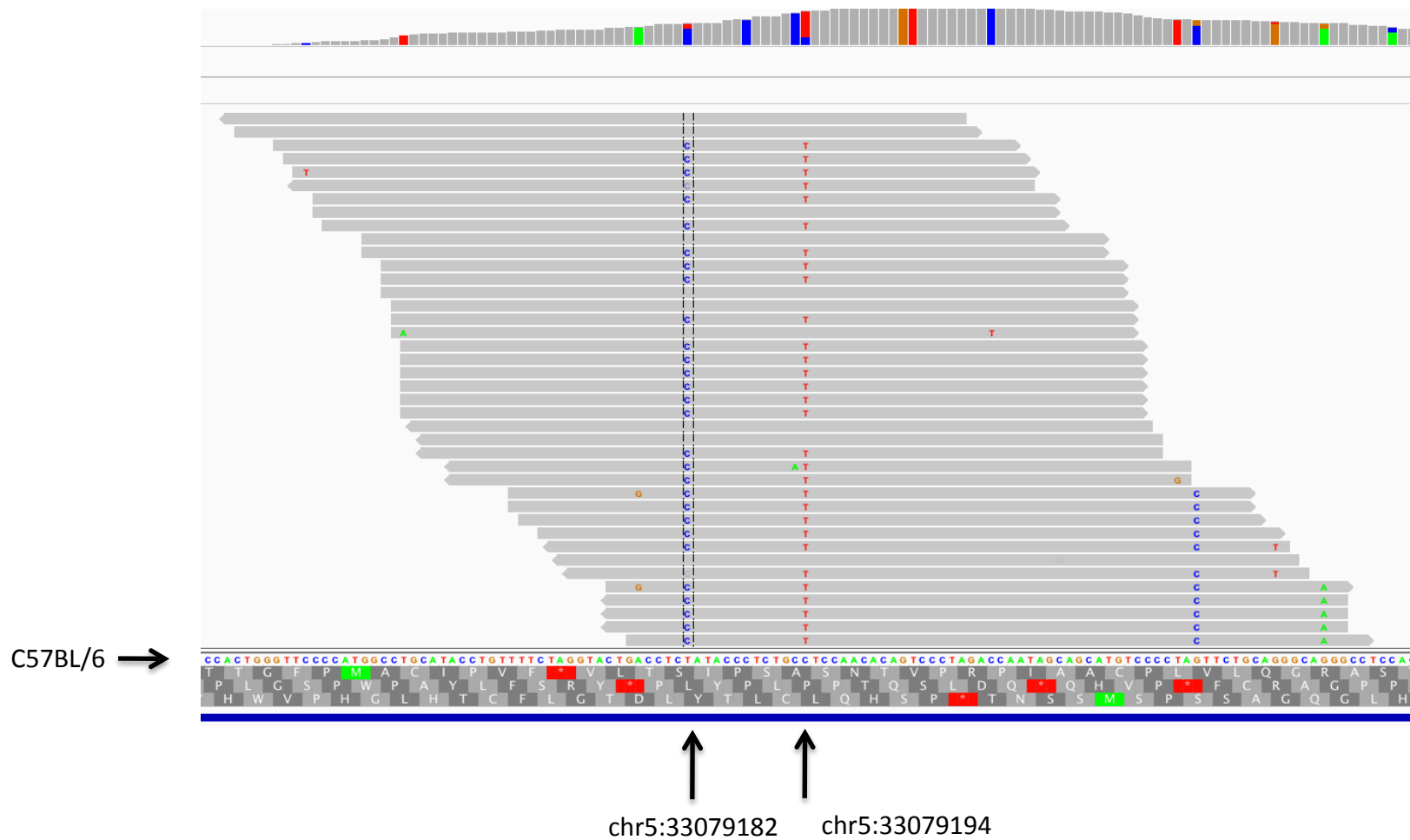
The read alignments shown are only for the forward strand. Considering that the reverse strand codes for Pisd, Chr5:33079182 T-to-C is actually a canonical A-to-G event. The C to T edit on the same forward mapped read shown here, Pisd 33079194, would be a non canonical G to A on the reverse coding strand.

Figure S2 B Pisd
 Sample 1: Male Chow DxB 3 mice pooled



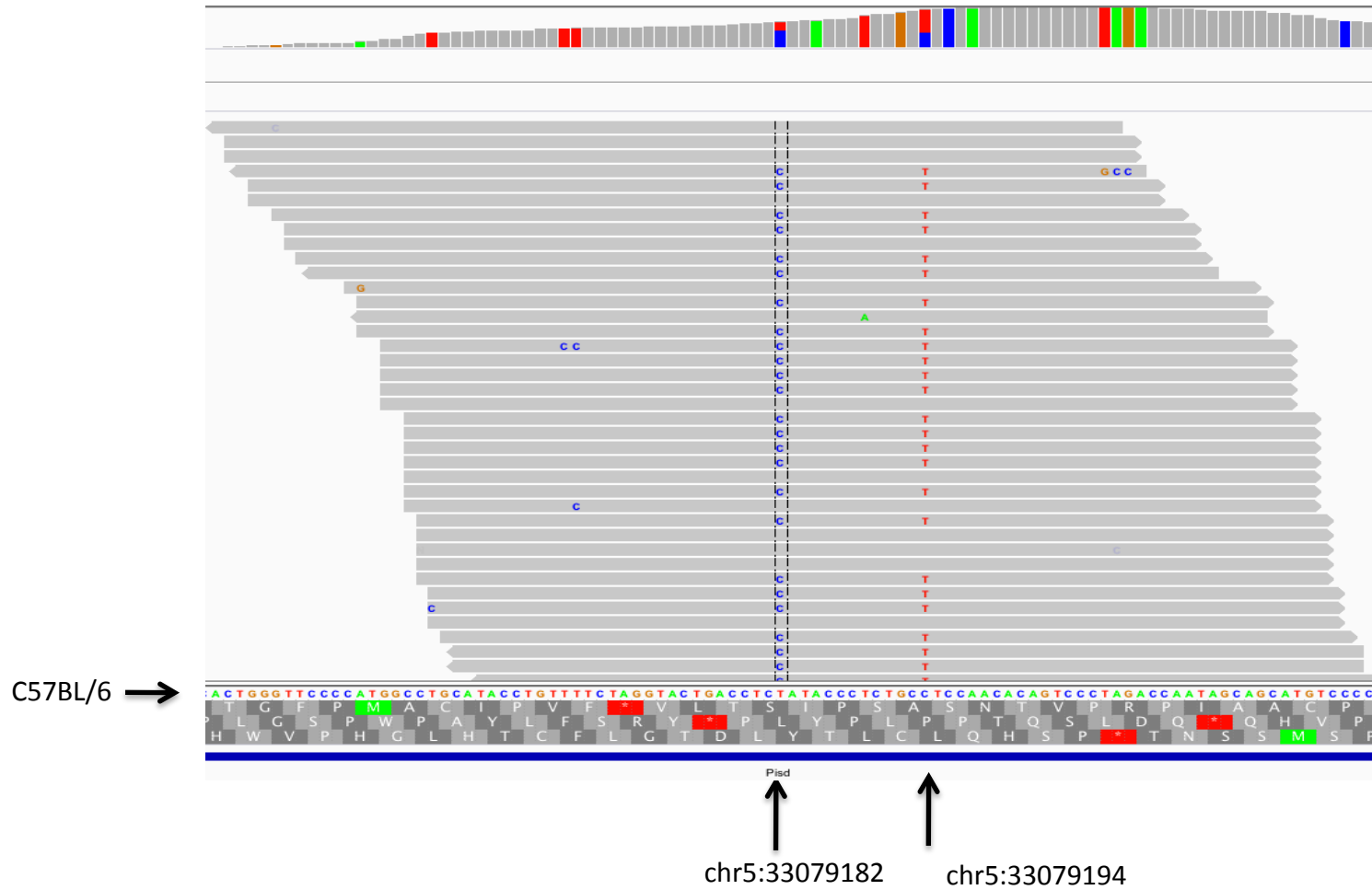
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 C Pisd
 Sample 2: Male High Fat Diet DxB 3 mice pooled



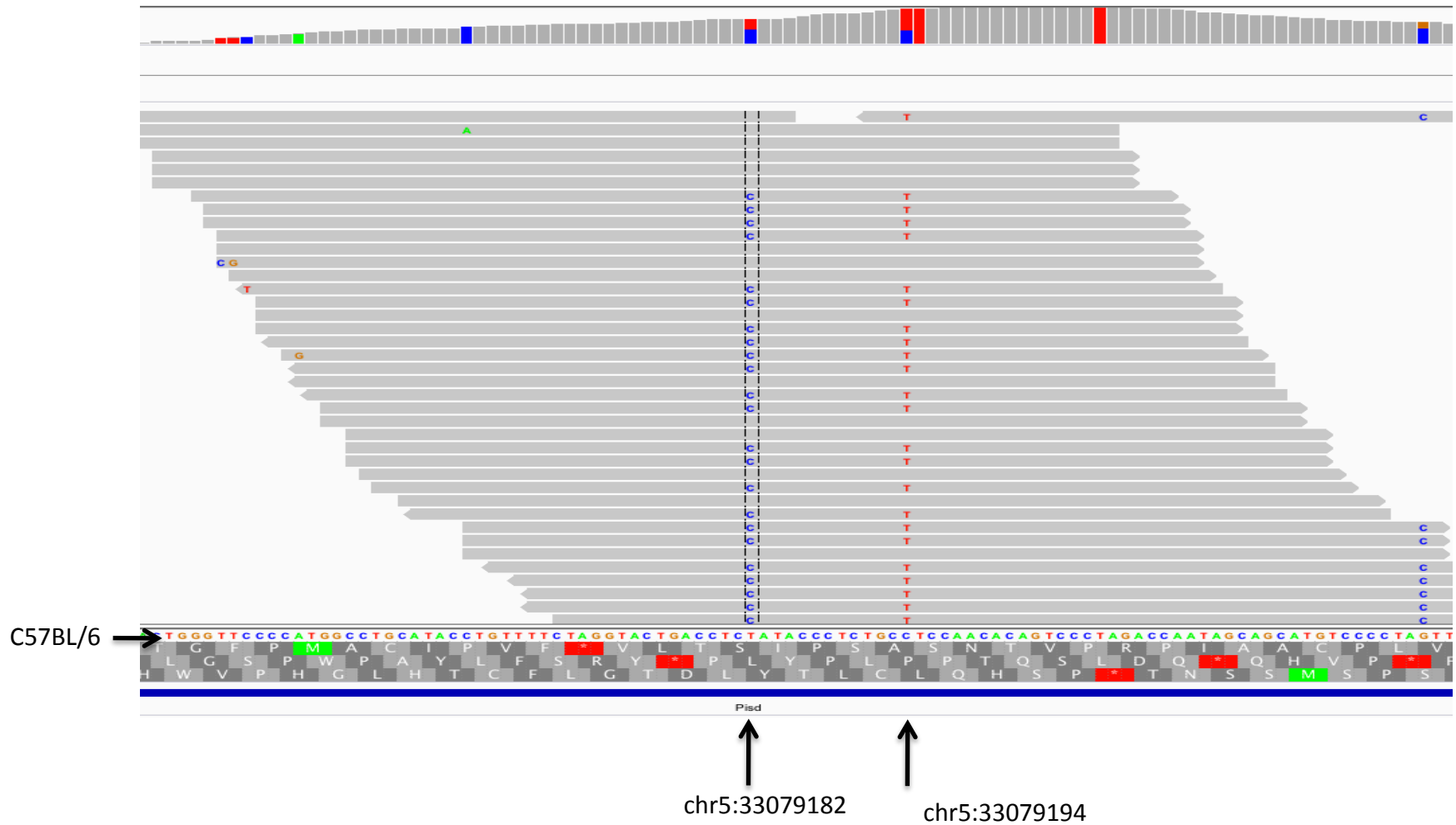
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 D Pisd
 Sample 5: Female High Fat Diet DxB 3 mice pooled



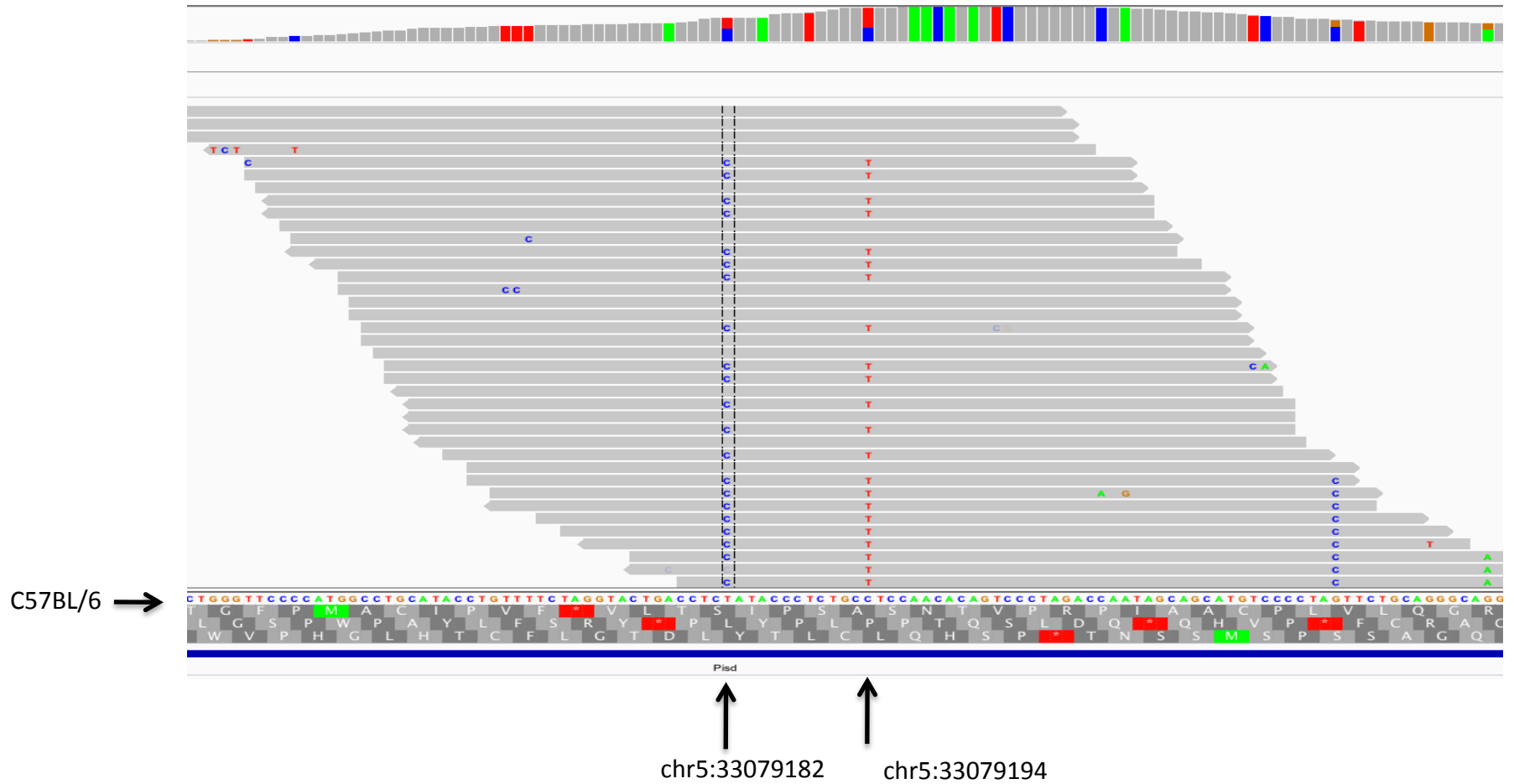
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 E Pisd
 Sample 6: Male Chow BxD 3 mice



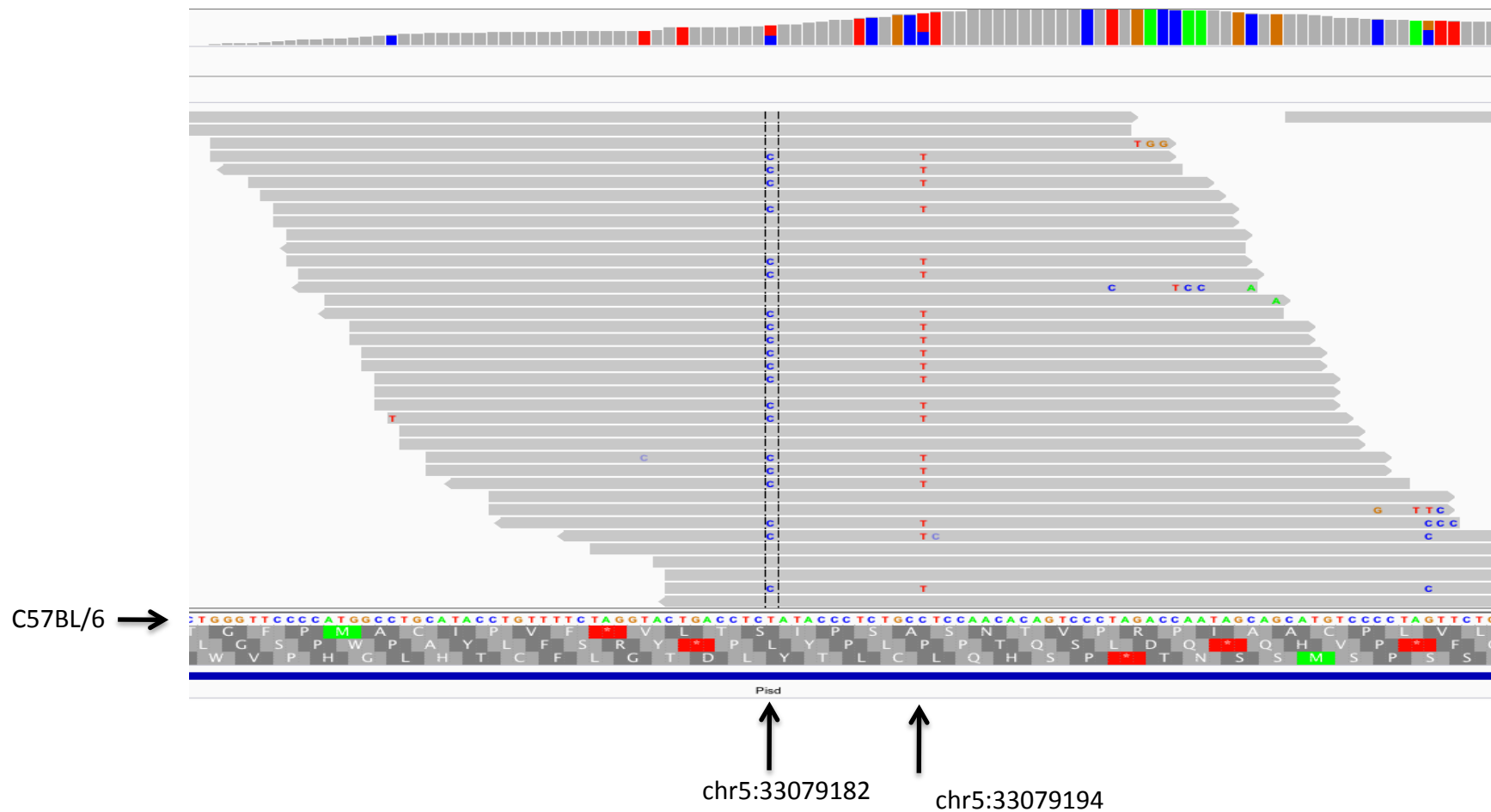
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 F Pisd
 Sample7: Male High Fat BxD 3 mice pooled



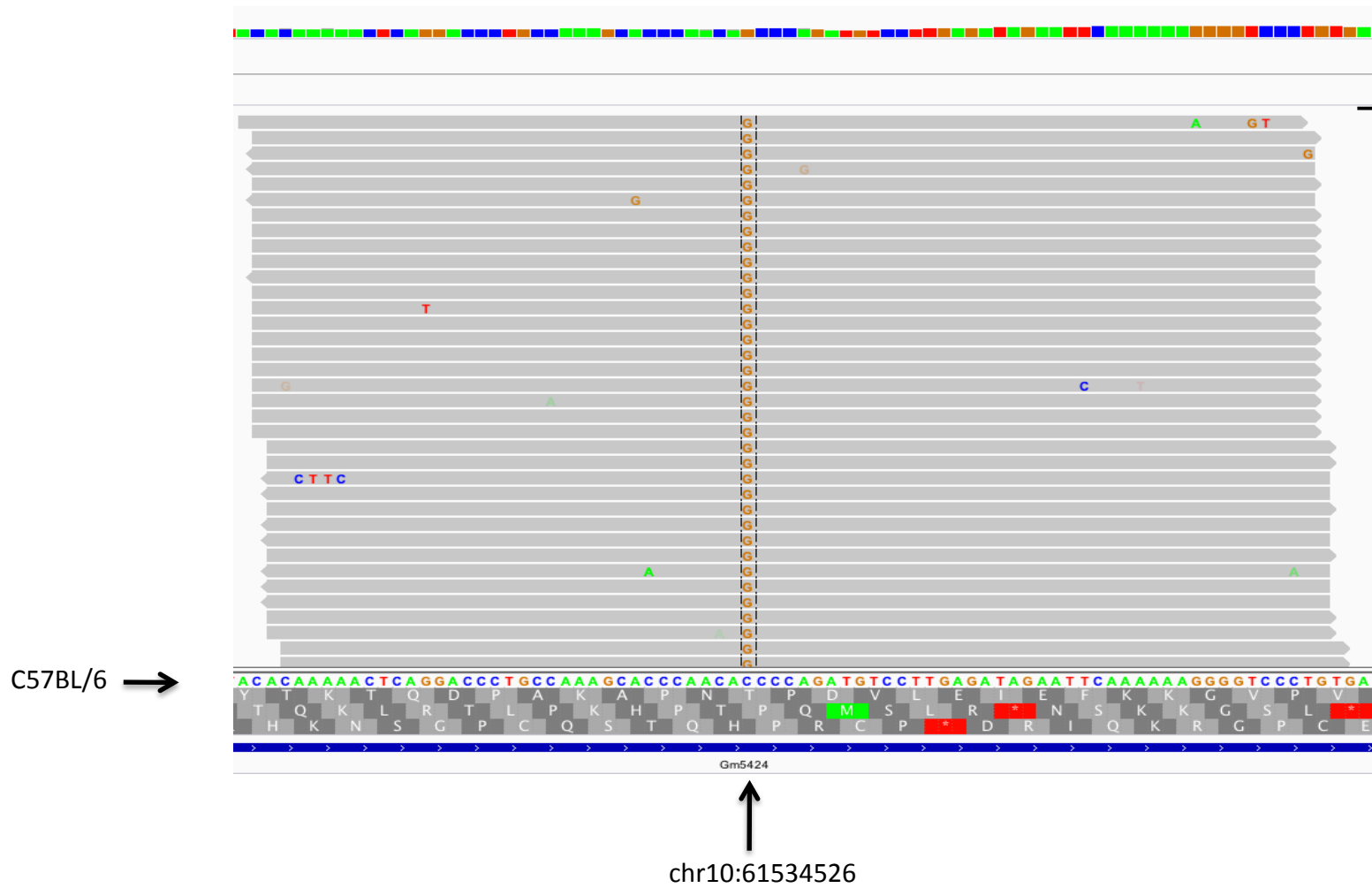
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 G Pisd
 Sample10: Female High Fat BxD 3 mice pooled



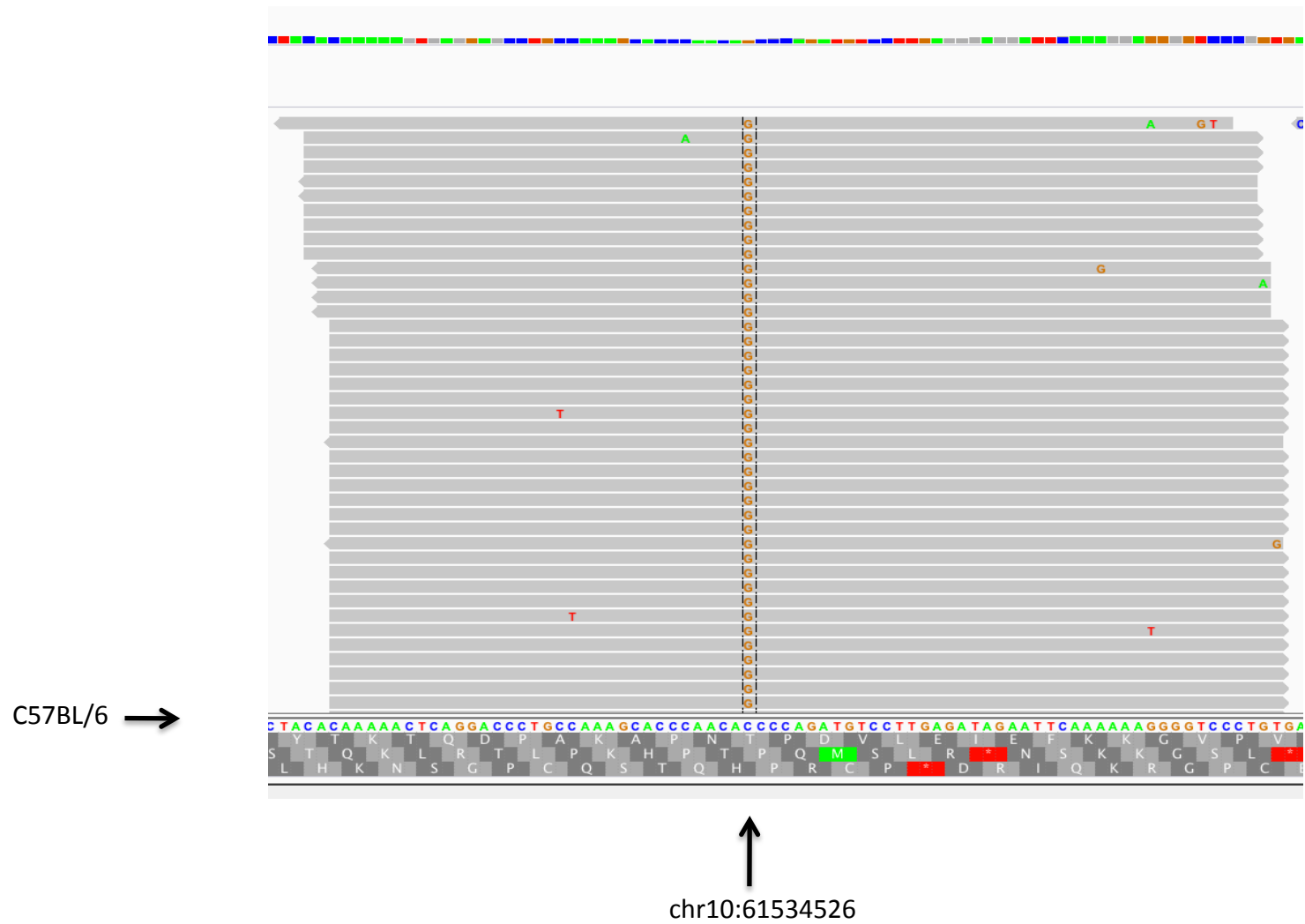
The IGV screenshot of Sample2 forward read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'T' at 9182 , however, most of the reads support the 'C' allele. Vice versa for 9194

Figure S2 H Gm5424
 Sample 1: Male Chow DxB 3 mice pooled



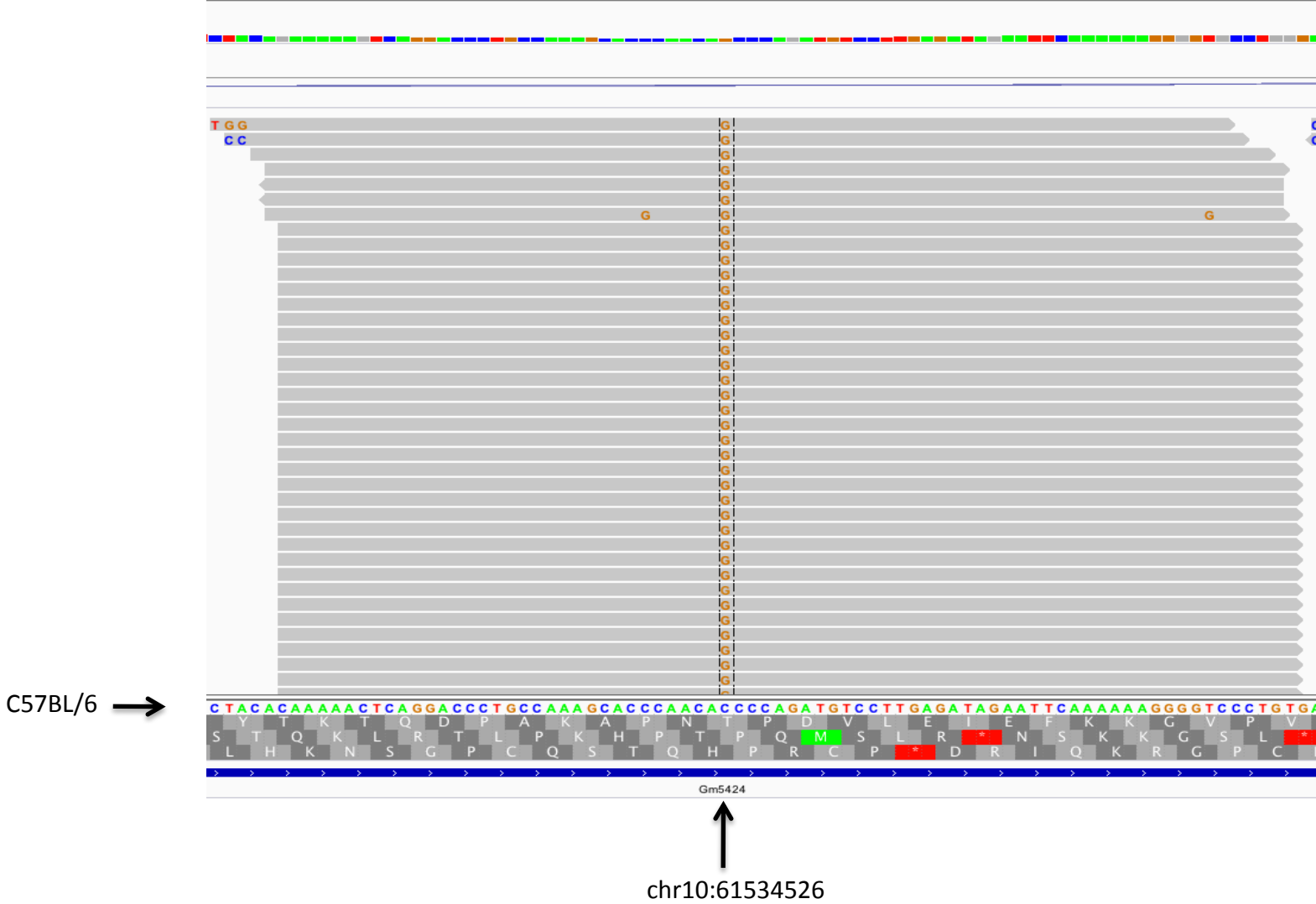
The IGV screenshot of Sample1 read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'C' , however, most of the reads support the 'G' allele.

Figure S2 I Gm5424
Sample 2: Male High Fat Diet DxB 3 mice pooled



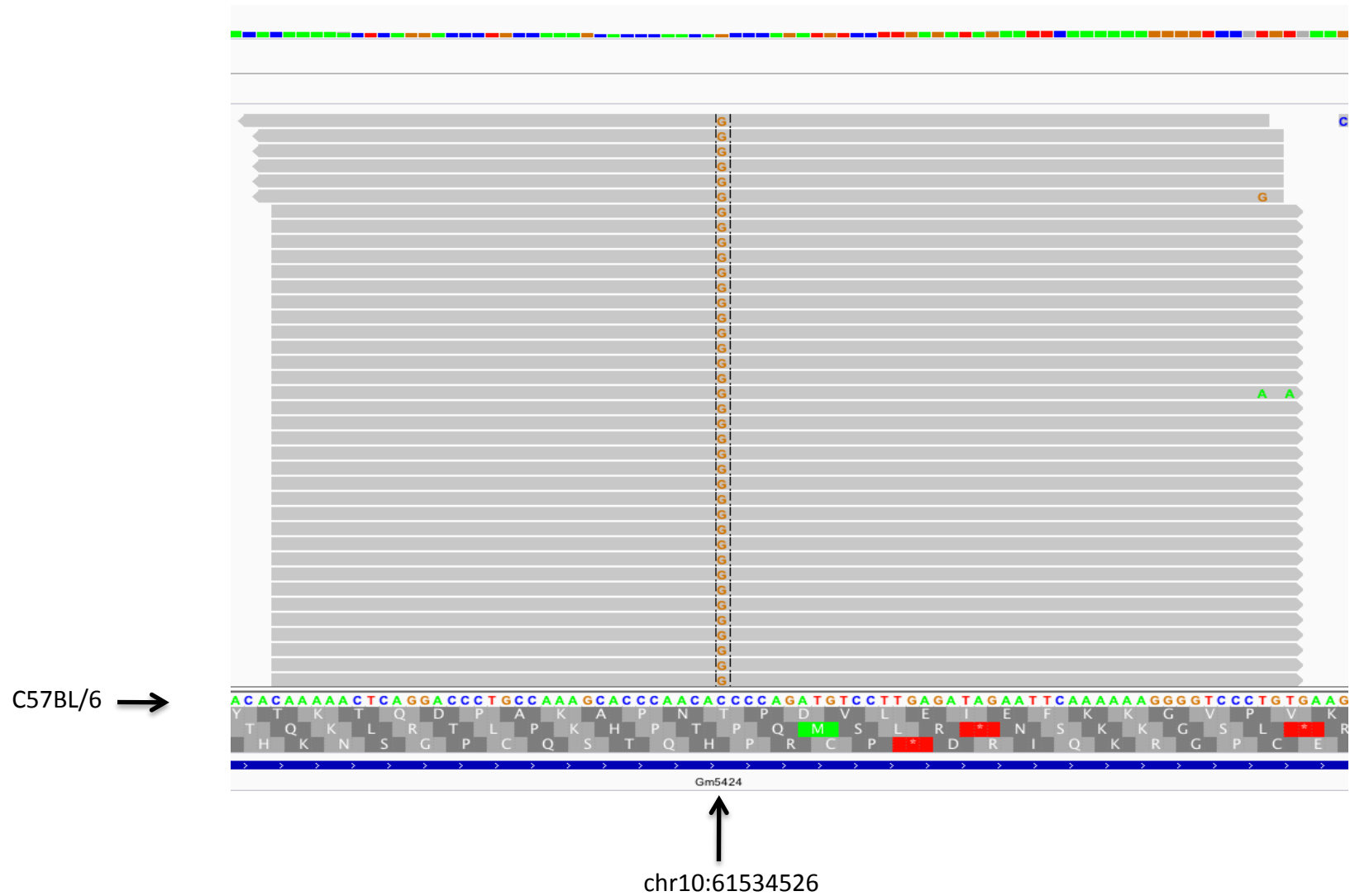
The IGV screenshot of Sample2 read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'C' , however, most of the reads support the 'G' allele.

Figure S2 J Gm5424
 Sample 5: Female High Fat Diet DxB 3 mice pooled



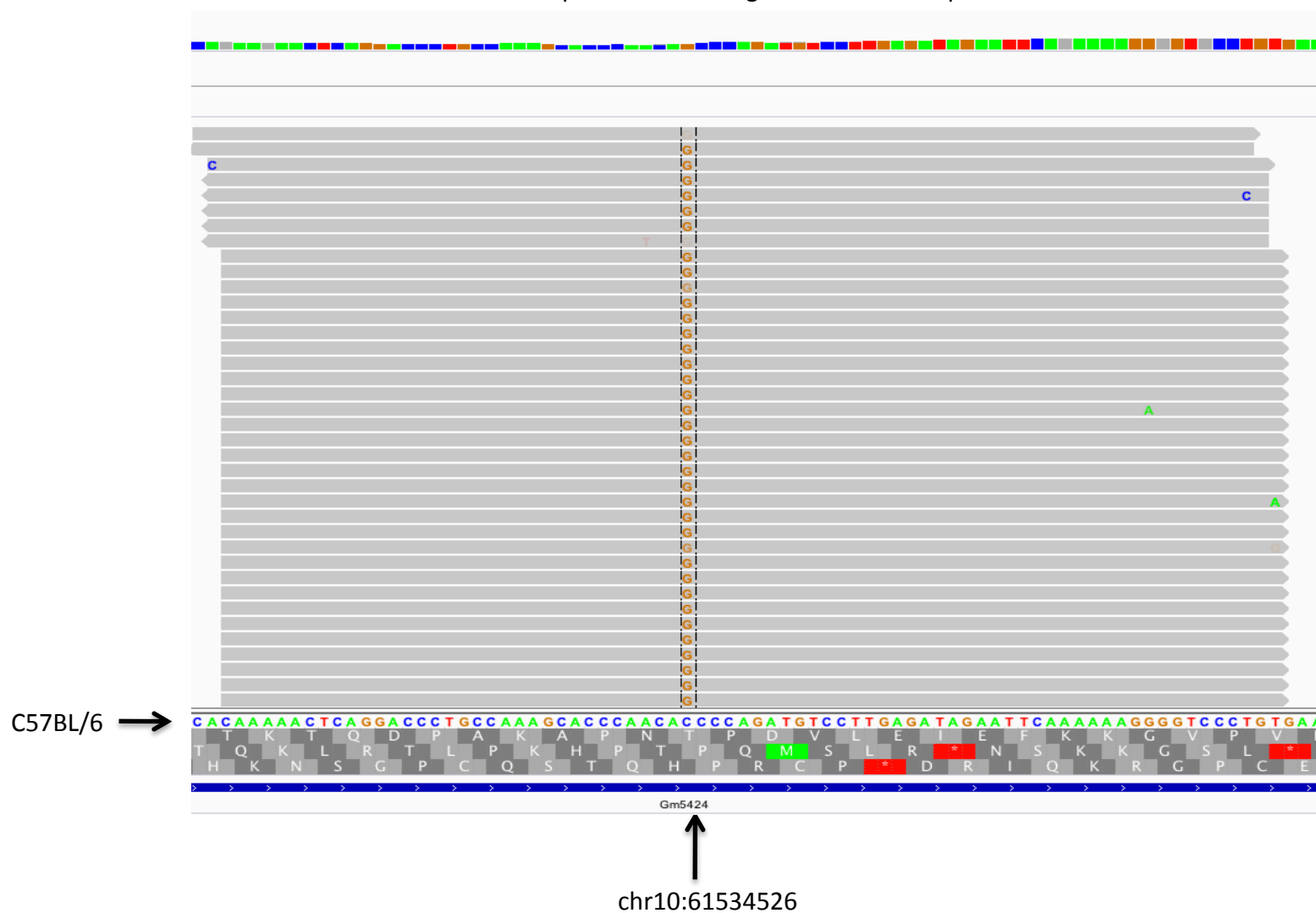
The IGV screenshot of Sample5 read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'C' , however, most of the reads support the 'G' allele.

Figure S2 L Gm5424
Sample7: Male High Fat BxD 3 mice pooled



The IGV screenshot of Sample7 read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'C' , however, most of the reads support the 'G' allele.

Figure S2 M Gm5424
 Sample10: Female High Fat BxD 3 mice pooled



The IGV screenshot of Sample10 read alignment. Each grey vertical line indicates a read. The reference(C57BL/6 build mm9) allele is 'C', however, most of the reads support the 'G' allele.

Gm5424_4526 C to G chr10 61534526

140 bp up and downstream from the edited base

AAGAGCCCCTGGAGTATGGATGAAAACCTCATGCACATCAGCTATGAGGCTGGGATCCTGGAAAACCCCAAGATCAAGCACCTCCGGGTCTCTACAAAAAAGCTCAGGACCTGCCAAAGCACCCAACA(C/G)CCCAGATGTCCTTGAG
ATAGAATCAAAAAGGGGTCCCTGTGAAGGTGACCAACATCAAGATGGCACAACCCGCACCACATCCCTGGAAGTCTTCATGTACCTGAACGAAGTTCGGGGCAAGCACGGAGTGGGTCGCATTGACATCGTGGAGA

blasted with this sequence from the above sequence
CTCAGGACCTGCCAAAGCACCCAACACCCAGATGTCCTTGAGATAGAA

Strand=Plus/Plus
**alignment of reference DNA and Sanger sequenced genomic DNA from BL/6 (Query)
showing base is homozygous at that location**

Query 78 GACCCTGCCAAAGCACCCAACA**C**CCAGATGTCCTTGAGATAGAA 122
|||||
Sbjct 6 GACCCTGCCAAAGCACCCAAC**C**CCAGATGTCCTTGAGATAGAA 50

Strand=Plus/Plus
**alignment of reference sequence (sbjct) and Sanger sequenced cDNA from BL/6 (query)
RDD is present**

Query 76 GACCCTGCCAAAGCACCCAACA**G**CCAGATGTCCTTGAGATAGAA 120
|||||
Sbjct 6 GACCCTGCCAAAGCACCCAAC**G**CCAGATGTCCTTGAGATAGAA 50

Strand=Plus/Plus
**alignment of reference DNA and Sanger sequenced genomic DNA from DBA (query)
showing the base is homozygous at that location**

Query 71 CTCAGGACCTGCCAAAGCACCCAACA**C**CCAGATGTCCTTGAGATAGAA 120
|||||
Sbjct 1 CTCAGGACCTGCCAAAGCACCCAAC**C**CCAGATGTCCTTGAGATAGAA 50

Strand=Plus/Plus
**alignment of reference sequence (sbjct) and Sanger sequenced cDNA from DBA (query)
RDD is present**

Query 69 GACCCTGCCAAAGCACCCAACA**G**CCAGATGTCCTTGAGATAGAA 118
|||||
Sbjct 6 GACCCTGCCAAAGCACCCAAC**G**CCAGATGTCCTTGAGATAGAA 50

Pisd A to G chr5 33079182, and Pisd G to A chr5 33079194

reference sequence 100 bp up and downstream from edited base

AAATTCATGGCATTCCAAGATCGACATGAACTCTGGTCTTGCCCTACCACTGGGTTCCCATGGCCTGCATACCTGTTTTCTAGGTACTGACCTC(T/C)ATACCCTCTGCCTCCAACACAGTCCCTAGACCAA
TAGCAGCATGTCCCCTAGTTCTGCAGGGCAGGGCTCCACCTGCCCTTCACTCCAGACTTACCA
TTTTCTAGGTACTGACCTC(T/C)ATACCCTCTGCCTCCAACACAG--using this smaller fragment to blast against Sanger sequences

Strand=Plus/Minus
Reverse primer Sanger sequence of BL/6 genomic DNA (Query) against reference

Query 88 CTGTGTTGGAGGCAGAGGGTATAGAGGTCAGTACCTAGAAAA 129
|||||
Sbjct 42 CTGTGTTGGAGGCAGAGGGTATAGAGGTCAGTACCTAGAAAA 1

Strand=Plus/Minus
Reverse primer Sanger sequence of DBA gDNA (Query) against the reference

Query 87 CTGTGTTGGAGGCAGAGGGTATAGAGGTCAGTACCTAGAAAA 128
|||||
Sbjct 42 CTGTGTTGGAGGCAGAGGGTATAGAGGTCAGTACCTAGAAAA 1

Strand=Plus/Minus

Reverse primer Sanger sequence of cDNA (Query) against BL/6 reference sequence

```

Query 43 CTGTGTTGGAAGCAGAGGGTATCGAGGTCAGTACCTAGAAAA 84
          |||
Sbjct 42 CTGTGTTGGAGGCAGAGGGTATAGAGGTCAGTACCTAGAAAA 1

```

Strand=Plus/Minus

Reverse primer Sanger sequence of cDNA (Query) from DBA against reference sequence

```

Query 43 CTGTGTTGGAAGCAGAGGGTATCGAG 68
          |||
Sbjct 42 CTGTGTTGGAGGCAGAGGGTATAGAG 16

```

shows A to G editing in red and G to A editing in blue

Strand=Plus/Plus

alignment of reference DNA and Sanger sequenced genomic DNA from BL/6 (Query) showing base is homozygous at that location

```

Query 218 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 259
          |||
Sbjct 1 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 42

```

Strand=Plus/Plus

alignment of reference DNA and Sanger sequenced genomic DNA from DBA (Query) showing the base is homozygous at that location

```

Query 216 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 257
          |||
Sbjct 1 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 42

```

Strand=Plus/Plus

Alignment of reference DNA with Sanger sequenced cDNA from BL/6 (Query) showing that the base is edited

```

Query 82 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 123
          |||
Sbjct 1 TTTTCTAGGTAAGTACTGACCTCTATACCTCTGCCTCCAACACAG 42

```

Strand=Plus/Plus

Alignment of reference DNA with Sanger sequenced cDNA from DBA (Query) showing that the base is edited

```

Query 89 CCTCCATACCTCTGCCTCCAACACAGTCCCTAGACCAATAGCAGCATGTCCCCTA 148
          |||
Sbjct 97 CCTCTATACCTCTGCCTCCAACACAGTCCCTAGACCAATAGCAGCATGTCCCCTA 156

```

This website was used to compare the Sanger sequences with the reference sequences

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=blast2seq

Figure S3: Alignment of Sanger sequences for three RDD *Gm5424* CG, 10:61534526, *Pisd* A-to-G 5:33079182 and *Pisd* G-to-A 5:33079194 providing additional confirmation for non-canonical C-to-G and G-to-A RDD.

Tables S1-S4

Supporting Tables

Available for download at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.149054/-/DC1>.

Table S1: Characteristics of the 6 liver and 4 adipose tissue samples subjected to RNA-Seq.

Table S2: List of the 293 RDD in the liver of which 63 RDDs were selected after different filters. "chr": chromosome; "rdd.position": coordinate on the genome; "rdd": RNA DNA difference, the first are equal to the reference DNA base bases and the second to the edited base; "startExon" and "endExon": coordinate of the start and the end of the exon using UCSC database (in which base 1 is coded "0"); "shared_by_adipose": RDD sites also observed in adipose data; "sequenom": indicates if the RDD sites have been analyzed by the sequenom technology and gives the results (validated versus not_validated); "splice.junction_bias" indicates if the RDD is nearby the extremity of the exon (less than 6 bases). The distance of the RDD position and the exon start is indicated in the column "rdd.position_startExon", the distance of the RDD position and the exon end is indicated in the column "endExon_rdd.position"; "Sequence.repeat_bias": Yes = RDD at proximity of a simple sequence repeat; "Unidirectional.strand_bias": Yes = RDD site for which the proportion of reads belonging to the forward sequencing strand is $\geq 20\%$ and $\leq 80\%$; « Read.extremity_bias »: Yes = RDD site for which the proportion of reads where the alternative base is found in the five first or last bases of the sequencing read is $\leq 50\%$; "DNABase_mean": mean of the read counts for the DNA base in the 4 or 6 different samples; "EditedBase_mean": mean of the read counts for the edited base in the 4 or 6 different samples; "ratio_EditedBase.mean_on_Total": EditedBase_mean / EditedBase_mean + DNABase_mean. "EST_DNA.base": number of EST found with the DNA base. "EST_Edited.base": number of EST found with the edited base; "ratio_EditedESTonTotalEST": EST_Edited.base / EST_Edited.base + EST_DNA.base. The 12 last columns correspond to the read count of the DNA base and edited base for the 6 hepatic samples.

Table S3: List of the 1667 RDD in adipose tissue in which 188 remain after different filters. Same column abbreviations as in Table S2.

Table S4: Details of Sequenom results for 7 RDD verifying the homozygosity of the DNA sequences and the proportion of total edited base sequences detected.