Supplemental Information

Mild inborn errors of metabolism in commonly used inbred mouse strains

João Leandro[1], Sara Violante[1,2], Carmen A. Argmann[1], Jacob Hagen[1], Tetyana Dodatko[1], Aaron Bender[1], Wei Zhang[2], Evan G. Williams[3], Alexis M. Bachmann[4], Johan Auwerx[4], Chunli Yu[1,2], Sander M. Houten[1]

[1]Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, Box 1498, New York, NY 10029, USA; [2]Mount Sinai Genomics, Inc, One Gustave L Levy Place #1497 NY, NY 10029; [3]Department of Biology, Institute of Molecular Systems Biology, ETH Zürich, Zürich CH-8093, Switzerland; [4]Laboratory of Integrative and Systems Physiology, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Corresponding author: Sander Houten, Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, Box 1498, New York, NY 10029, USA. Phone: +1 212 659 9222, Fax: +1 212 659 8754, E-mail: sander.houten@mssm.edu

Extended Results (related to the molecular characterization of the variants)

Identification and characterization of the Dhtkd1 variant causing α-aminoadipic acidemia in mice
Data available through the Mouse Genomes Project list several SNPs, insertions/deletions and structural variants in the *Dhtkd1* locus [1, 2]. We focused our studies on one structural variant, a 1042 base pair (bp) deletion in intron 4 (chr2:5,924,441-5,925,482) reported in all strains known to express high level of DHTKD1 [3, 4]. This intronic region is also of particular interest because it was found in two expressed sequence tag (EST) clones (AA414856 and CD803297). The sequences of these ESTs suggest that this region in the *Dhtkd1* locus of C57BL/6J mice contains an alternate exon with splice donor and acceptor sites. The open reading frame of a hypothetical mRNA containing this additional exon is out of frame and encodes a truncated protein. Nonsense-mediated decay of this mRNA would offer an explanation for the decreased *Dhtkd1* mRNA expression in C57BL/6J mice. We then investigated whether the predicted alternate splice donor site in C57BL/6J mice is used and PCR amplified the predicted alternate mRNA in liver cDNA. Sanger sequencing of this product confirmed that the alternate splice donor site was located at position chr2:5,924,677.

Upon closer inspection, the mapped read data available through the Mouse Genomes Project appear more consistent with a smaller deletion. In order to determine the exact size of this deletion, we amplified and Sanger sequenced this genomic region in C57BL/6J, DBA/2J and 129S2/SvPasCrl. Our data reveal the presence of a 556 bp deletion in DBA/2J and 129S2/SvPasCrl (chr2:5,924,595-5,925,150). The distribution of this structural variant in different inbred mouse strains based on data from the Mouse Genomes Project is displayed in Figure S1A. This deleted region contains the alternate splice donor site present in C57BL/6J. In the GRCm38 genome assembly, this region is annotated as a repeat region called MT2_Mm. This sequence is 86% identical to the long terminal repeat (LTR) from the mouse endogenous retroviral (ERV) sequence MuERV-L (Y12713.1), a class III endogenous retrovirus [5, 6]. Taken together, these data suggest that the DHTKD1 defect in C57BL/6J may be caused by the presence of a solitary LTR harboring an alternative splice donor site.

RNA sequencing reveals the splicing defect associated with the Dhtkd1^{B6} allele
We next used a B6129PF2 cohort of mice to investigate the molecular mechanism underlying DHTKD1 deficiency associated with the *Dhtkd1^{B6}* allele. We first genotyped the *Dhtkd1* locus in these mice using two different methods: direct amplification of the locus around the structural variant and a restriction fragment length polymorphism (RFLP) assay for rs27092974. We were

able to differentiate mice carrying the B6 allele from mice carrying the 129 allele (Figure S1B). The SNP and structural variant are in complete linkage disequilibrium as expected based on the BXD mapping data [3]. In order to further study the alternate splicing event, we visualized the aligned reads for the *Dhtkd1* region of interest and grouped the animals according to their genotype; *Dhtkd1*$^{B6/B6}$, *Dhtkd1*$^{B6/129}$ or *Dhtkd1*$^{129/129}$. In *Dhtkd1*$^{B6/B6}$ animals, we observed many intronic reads covering intron 4 from exon 4 to the alternate splice donor site at position chr2:5,924,677. *Dhtkd1*$^{129/129}$ animals had almost no intronic reads, whereas *Dhtkd1*$^{B6/129}$ had intermediate levels (Figure S1C). This indicates that the alternate splice donor in intron 4 competes with the splice donor at exon 4. We also displayed the number of junction reads that join exons. *Dhtkd1*$^{B6/B6}$ animals showed joining of intron 4 to exon 5 at the alternate splice donor site at position chr2:5924677, which appeared to be at the expense of joining of exon 4 to exon 5. For all these observations there was a clear gene dosage effect (Figure S1C). The same observations were made upon comparison of liver RNA sequencing data generated from C57BL/6J and DBA/2J mice (not shown).

We next used linear regression to estimate the effect size of the deletion allele (Figure S2A). We first counted the number of reads from the splice donor site of exon 4 to the alternate splice donor site in intron 4. We found 14.6 intronic reads per million (RPM) in *Dhtkd1*$^{B6/B6}$ animals, which was reduced by 6.7 RPM for each copy of the 129 deletion allele (Figure S2A). Total exonic *Dhtkd1* reads increased by 21 RPM for each copy of the 129 deletion allele, which indicates that the alternatively spliced *Dhtkd1* mRNA is less stable and may be subjected to nonsense-mediated decay (Figure S2A).

As a consequence of the alternate splicing, total exonic reads are not an accurate parameter for estimating the change in functional *Dhtkd1* mRNA molecules. This is because the alternatively spliced intron 4-containing mRNAs may contain all exons, but do not encode a functional protein. In order to estimate the effect size of the deletion on the levels of functional *Dhtdk1* mRNA, we counted only junction reads joining exon 4 and exon 5, which are derived from *Dhtkd1* mRNAs that escaped the alternate splicing event. These reads, which reflect functional *Dhtkd1*, increased by 38% for each copy of the 129 deletion allele (Figure S2A). Thus *Dhtkd1*$^{B6/B6}$ animals have a 4.2-fold decrease in functional *Dhtdk1* when compared to *Dhtkd1*$^{129/129}$ animals.

Experimental confirmation of the biochemical defect associated with Dhtkd1^{B6}

We next sought to confirm the data obtained by RNA sequencing. We used quantitative PCR and amplified *Dhtdk1* using a primer set within exon 5 and a primer set amplifying from exon 4

to exon 5 (Figure S2B). While the former will amplify nonfunctional and functional mRNA, the latter primer set only detects functional *Dhtdk1* mRNA. The qPCR confirms that *Dhtkd1*[B6/B6] animals have the lowest *Dhtkd1* expression with intermediate levels in *Dhtkd1*[B6/129] animals and the highest levels in *Dhtkd1*[129/129] animals. The decrease in *Dhtkd1* mRNA in *Dhtkd1*[B6/B6] animals was more pronounced for functional *Dhtkd1* mRNA (-58%) when compared to total *Dhtkd1* mRNA (-43%), but the slopes were not significantly different due to unexplained variation in the *Dhtkd1*[129/129] animals (Figure S2B). The difference between functional and total *Dhtkd1* mRNA is more consistently captured by the ratio of functional over total *Dhtkd1* mRNA (Figure S2B). This showed that in addition to an overall decrease in total *Dhtkd1* mRNA in *Dhtkd1*[B6/B6] animals, only 73% of this *Dhtkd1* mRNA is functional. These results are consistent with the RNA sequencing data.

Identification of the causal variant for the branched-chain ketoacid dehydrogenase defect in C57BL/6J mice

We next set out to identify the molecular mechanism and causal variant underlying BCKDHB deficiency in C57BL/6J mice [3, 7]. The Mouse Genomes Project lists one large structural variant that is common to 129, DBA/2J and all other tested strains, a deletion of 7332 bp in intron 1 (Chr9:83,942,547-83,949,878). In the C57BL/6J reference genome this region is an LTR retrotransposon (IAPLTR1_Mm, Figure S3A). Two *Bckdhb* transcript variants are described; the canonical variant (NM_001305935.1) and a variant containing an alternate exon 1 encoding a protein that uses a potential downstream startcodon (Met69, NM_199195.1). This alternate exon 1 is part of the LTR retrotransposon in intron 1 (chr9:83,948,781-83,948,915) and the encoded variant protein lacks the mitochondrial targeting sequence; therefore, it is unlikely to be a functional part of the BCKDH complex.

In order to study the molecular mechanism underlying the decreased *Bckdhb* expression in C57BL/6J mice, we used the RNA sequencing data of the B6129PF2 mouse livers. Mice were grouped as *Bckdhb*[B6/B6], *Bckdhb*[B6/129] or *Bckdhb*[129/129] based on the genotype for rs13480313 (reference = T; 129 = A; exon 11/11). In mice with the *Bckdhb*[B6/129] genotype, we found an allelic imbalance for the expression of this SNP (B6/129 = 0.70, Figure S3B). Total exonic *Bckdhb* reads were decreased by ~16% for each *Bckdhb*[B6] allele (Figure S3C). We next visualized reads and splice junctions focusing on the first exons (Figure S3A). In mice with one or two *Bckdhb*[B6] alleles, we observed; (1) usage of alternate exon 1 and joining with exon 2 (Figure S3D), (2) decreased joining of exon 1 with exon 2 (Figure S3E), (3) skipping of exons 2 and 3 (Figure S3F), (4) various other abnormal splicing events, and (5) transcriptional activity in

4

intron 1. There was a clear gene dosage effect for all these events. None of these abnormalities were observed in mice with the *Bckdhb*^{129/129} genotype. The same observations were made upon comparison of liver RNA sequencing data generated from C57BL/6J and DBA/2J mice (not shown).

Total exonic reads is probably not the best estimate for the degree of BCKDHB deficiency. In order to estimate the effect size of the *Bckdhb*^{B6} allele on the BCKDHB function, we counted reads joining exon 1 with exon 2, which is most likely a better estimate for functional *Bckdhb* mRNA level. Reads joining exon 1 with exon 2 decreased by 30% for each copy of the *Bckdhb*^{B6} allele (Figure S3E). If we assume that reads joining exon 1 with exon 2 represent functional mRNAs and the allelic imbalance at rs13480313 reflects nonsense-mediated decay of all nonfunctional mRNAs, then the residual BCKDHB protein expression in C57BL/6J is approximately 35% to 40% when compared to strains without the structural variant. Immunoblotting results demonstrated that the actual residual BCKDHB protein expression is ~2-fold lower. We estimated that the residual BCKDHB protein expression is 18% in C57BL/6J (compared to 129S2/SvPasCrl) and 18% *Bckdhb*^{B6/B6} mice of the B6129PF2 cohort (Figure S3G).

Isovaleryl-CoA dehydrogenase deficiency in 129 mice

In order to identify a cause for the elevated isovalerylcarnitine in mice of the 129 substrains, we studied expression patterns of *Ivd* in the RNA sequencing data of the liver samples of B6129PF2 mice by grouping mice as *Ivd*^{B6/B6}, *Ivd*^{B6/129} or *Ivd*^{129/129} based on the genotypes for rs4905023 and rs47004510 in exon 9. We observed partial skipping of exon 10 in mice with *Ivd*^{B6/129} and *Ivd*^{129/129} genotypes (Figure S4A). Indeed, the number of reads joining exons 9 and 11 increased for each *Ivd*¹²⁹ allele (Figure S4B). Total *Ivd* reads were not affected by *Ivd* genotype (Figure S4C). No abnormalities were observed for *Acadsb* expression (not shown).

Next, we studied *Ivd* variants in 129 mice in order to identify the mutation causing this splice defect. We noted one synonymous SNP at the last nucleotide of exon 10 (rs27440099, K356) that changes the conserved G of a splice donor site into an A. This variant is therefore likely to decrease splicing efficiency and can be classified as an exonic splice mutation. This SNP is found in 129 (129P2/OlaHsd, 129S1/SvImJ and 129S5/SvEvBrd), BTBR *T*⁺ *Itpr3*^{tf}/J, BUB/BnJ and LP/J strains. In mice with the *Ivd*^{B6/129} genotype, we found an allelic imbalance for the expression of this SNP (129/B6 = 0.28, Figure S4D). PCR on cDNA further confirmed the aberrant splicing in mice *Ivd*^{B6/129} and *Ivd*^{129/129} genotypes (Figure S4E).

Supplemental Materials and Methods

Mouse genomic studies

A query for SNPs, indels, and structural variants in the *Dhtkd1*, *Bckdhb* and *Ivd* loci was performed using the Mouse Genomes Project website [1, 2]. The *Dhtkd1* locus was genotyped using two different methods. The locus around the structural variant was amplified using the following forward and reverse primers; 5'-AAA ACA TTG CCA CGA AGG AC-3' and 5'-AGG CAC CCA CAG AAT CCA TA-3'. This yields a product of 767 bp for *Dhtkd1*^{B6}, and 211 bp for *Dhtkd1*^{129} and *Dhtkd1*^{D2}. These products were Sanger sequenced to determine the size to the structural variant (GenBank: MH925964-MH925966). We also performed a restriction fragment length polymorphism (RFLP) assay for rs27092974 in exon 8 of *Dhtkd1*, which is approximately 6.7 kb downstream of the structural variant. For this we amplified 206bp of exon 8 containing rs27092974 using the following forward and reverse primers; 5'-ACC TAC GCA GAG CAC CTC AT-3' and 5'-TCT AGA GGC ACC CCT GTG TC-3'. The SNP introduces an SphI site and the PCR product is cleaved for *Dhtkd1*^{129} and *Dhtkd1*^{D2}, but not for *Dhtkd1*^{B6}. The *Bckdhb* and *Ivd* loci were genotyped using RNA sequencing data covering rs13480313 (*Bckdhb* exon 11), and rs4905023 and rs47004510 (both in exon 9 of *Ivd*).

Animal experiments

All animal experiments were approved by the IACUC of the Icahn School of Medicine at Mount Sinai or the local animal experimentation committee of the Canton de Vaud (Switzerland, license no. 2257.2), and comply with the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978). Six week old male 129S2/SvPasCrl, DBA/2J and C57BL/6J mice were ordered. At 8 weeks of age, blood for biochemical analyses was collected via a submandibular blood draw in the random fed state (4pm). Plasma was collected and stored at -20°C for future analyses. Mice were euthanized at 10 weeks of age exactly 2 hours after an intraperitoneal (ip) dose of L-lysine (100mg/kg). Immediately preceding necropsy, mice were anesthetized with pentobarbital (100mg/kg ip) and then exsanguinated via the vena cava inferior. Blood was collected for the preparation of EDTA plasma and organs were snap frozen in liquid nitrogen and stored at -80°C for future analyses.

In order to further study the molecular mechanisms underlying the DHTKD1, BCKDHB and IVD defects, we reanalyzed liver RNA sequencing data that were generated for a different project. In this project, we used *Ehhadh* (also known as multifunctional protein-1 or L-bifunctional protein) KO mice (B6;129P2-*Ehhadh*^{tm1Jkr}) [8] that were generously provided by Dr.

Janardan K Reddy (Department of Pathology, Feinberg School of Medicine, Northwestern University, Chicago, IL). *Ehhadh* KO mice were crossed with C57BL/6N in order to create a first generation progeny. The *Ehhadh* KO mouse was generated using BK4 ES cells [8], which are derived from the 129P2/OlaHsd strain and therefore the F1 contains genetic material of both the C57BL/6N and 129P2/OlaHsd strains (B6129PF1). The *Ehhadh*[+/-] mice from the F1 were intercrossed in order to generate an experimental cohort of B6129PF2 mice. From this F2 cohort, WT and *Ehhadh* KO mice were selected. Mice were euthanized at 29-45 weeks of age (mean age is 37 weeks) after overnight food withdrawal. Mice in group 1 had received an ip saline injection prior to food withdrawal. Plasma α-aminoadipic acid (Figure 1F) was also analyzed in mice from group 2. The mice in this group had received an ip injection of L-aminocarnitine (16mg/kg in saline). Before the necropsy, the mice were first anesthetized with pentobarbital (100mg/kg ip) and then exsanguinated via the vena cava inferior. Blood was collected for the preparation of EDTA plasma and organs were snap frozen in liquid nitrogen and stored at -80°C for further use. Livers were used for RNA sequencing analysis.

We show that this B6129PF2 dataset can be repurposed to further study the molecular mechanisms underlying the DHTKD1, BCKDHB and IVD defects in the C57BL/6 and 129 strains. Firstly, the deletion of *Ehhadh* had an overall small impact on the transcriptome, which is consistent with the observation that *Ehhadh* KO mice exhibit no detectable gross phenotypic defects [8]. Secondly, *Dhtkd1, Bckdhb* and *Ivd* were not differentially expressed between WT and *Ehhadh* KO animals. Thirdly, *Ehhadh, Dhtkd1, Bckdhb* segregated independently in the B6129PF2 cohort.

Male A/J, CAST/EiJ, NZO/HILtJ and PWK/PhJ mice were euthanized at 9-11 weeks old (mean age is 10 weeks). They were first anesthetized with a ketamine/xylazine mixture after which blood was collected by cardiac puncture. Plasma was prepared in Li-heparin tubes.

For the amino acid challenge experiments, 5 to 6 week old male 129S2/SvPasCrl, DBA/2J and C57BL/6J mice were ordered. At 9 weeks of age, all mice received an intraperitoneal injection of 15mL/kg 0.9% saline. At 12 weeks of age, all mice received 500mg/kg BCAAs (V150, L221, I129). At 14 weeks of age, all mice received 500mg/kg L-lysine. At 15 weeks of age, the BCAA challenge was repeated in the 129S2/SvPasCrl and C57BL/6J mice. One hour after the ip injection, a blood sample was collected from the saphenous vein using a capillary blood tube (Microvette CB 300 K2E).

mRNA expression studies

RNA was isolated using QIAzol lysis reagent followed by purification using the RNeasy kit (Qiagen). cDNA synthesis was performed with the SuperScript IV first strand synthesis system (Thermo Fisher Scientific) using random hexamers as primers.

The predicted alternate splice donor site in *Dhtkd1*[B6] was identified by PCR on C57BL/6J liver cDNA using the following forward and reverse primers; 5'-CTC TTG CTC TTG CTC CTG CT-3' and 5'-AGC ATT GTC ACA TGG AGG GG-3'. Sanger sequencing of the product confirmed that the alternate splice donor site was located at position chr2:5,924,677 as suggested by EST AA414856. Skipping of exon 10 associated with the *Ivd*[129] locus was identified by PCR on liver cDNA from B6129PF2 mice using the following forward and reverse primers; 5'-CCA CAC CAT TCC CTA CTT GC-3' and 5'-CTC GGC TGC ATA CAG AAT CA-3'.

Quantitative PCR was performed using the ABI Prism 7900HT with Bio-rad iQ SYBR Green Mastermix. Total *Dhtkd1* mRNA was quantified using forward and reverse primers in exon 5; 5'-GTG ATG TCC TGT CCC ACC TG-3' and 5'-CCA TCT TCC CGA GAC TGC TG-3'. Functional *Dhtkd1* mRNA was quantified using a forward primer in exon 4 and reverse primer in exon 5; 5'-GTT TGC CAC GGT GAA GAG AT-3' and 5'-TTA AGC CTC GCA TTT TAC GG-3'. *Bckdhb* was quantified using a forward primer in exon 2 and reverse primer in exon 3; 5'-ACA AGT GCC CTG GAT AAC TCA-3' and 5'-GCA TCG GAA GAC TCC ACC AA-3' Expression values were normalized using *Rplp0* (5'-ATG GGT ACA AGC GCG TCC TG-3' and 5'-GCC TTG ACC TTT TCA GTA AG-3'). All samples were analyzed in duplicate. Data were analyzed using Applied Biosystems SDS software.

RNA sequencing analysis

RNA samples were submitted to the Genomics Core Facility at the Icahn Institute and Department of Genetics and Genomic Sciences. cDNA libraries were prepared using the Illumina TruSeq RNA Library Preparation kit (#RS-122-2001). Samples were run on Illumina HiSeq 2500, and 100 nucleotide, single end fragments were read at a depth of about 25 million reads per sample. Raw and processed data were returned, count files were generated by aligning the reads to the mouse genome mm10 (GRCm38.75) with STAR and counting overlaps with exons grouped at gene level by featureCounts [9]. The aligned reads and splice junctions were visualized using the Integrative Genomics Viewer (IGV) and its Sashimi plot feature [10, 11].

Immunoblotting

Total liver protein homogenates were separated on a Bolt™ 4-12% Bis-Tris Plus Gel and blotted onto nitrocellulose. DHTKD1 was detected using a polyclonal antibody against recombinant human DHTKD1 (GeneTex, GTX32561). BCKDHB was detected using a polyclonal antibody against recombinant human BCKDHB (Abcam, ab201225). IVD was detected using a polyclonal antibody against recombinant human IVD (ThermoFisher, PA5-59547). Citrate synthase was detected using a polyclonal antibody against recombinant human protein (GeneTex, GTX110624). Secondary antibodies goat anti-rabbit IRDye 800CW or 680RD were from LI-COR Biosciences (Lincoln, NE, USA) and immunoblot images were obtained using the Odyssey infrared imaging system (LI-COR Biosciences).

Metabolite analysis

EDTA plasma (8 week old, random fed) was shipped to the Stable Isotope & Metabolomics Core of the Einstein-Mount Sinai Diabetes Research Center for a targeted metabolomics assay using the AbsoluteIDQ p180 kit (BIOCRATES Life Sciences AG, Innsbruck, Austria). This method quantifies ~180 metabolites in plasma including α-aminoadipic acid using liquid chromatography and flow injection analysis–mass spectrometry.

Plasma acylcarnitines, plasma amino acids and urine organic acids were measured by the Mount Sinai Biochemical Genetic Testing Lab (now Sema4). Plasma amino acids were analyzed as described [12]. Urine organic acids were quantified using a standard curve and pentadecanoic acid as internal standard.

Statistics

Data are displayed as the mean ± the standard deviation (SD) as indicated in the figure legends. Differences between groups of mice were evaluated using a two-sided t-test, one-way analysis of variance with Dunnett's multiple comparison test, a Kruskal-Wallis test or a two-way analysis of variance as indicated. Significance is indicated in the figures. Allele effect sizes were determined using linear regression analysis. All analyses were performed in GraphPad Prism 6.

References

[1]   B. Yalcin, K. Wong, A. Agam, M. Goodson, T.M. Keane, X. Gan, C. Nellaker, L. Goodstadt, J. Nicod, A. Bhomra, P. Hernandez-Pliego, H. Whitley, J. Cleak, R. Dutton, D. Janowitz, R. Mott, D.J. Adams, J. Flint, Sequence-based characterization of structural variation in the mouse genome Nature 477 (2011) 326-329.

[2]   T.M. Keane, L. Goodstadt, P. Danecek, M.A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N.A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T.G. Belgard, P.L. Oliver, R.E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C.A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I.J. Jackson, A. Czechanski, J.A. Guerra-Assuncao, L.R. Donahue, L.G. Reinholdt, B.A. Payseur, C.P. Ponting, E. Birney, J. Flint, D.J. Adams, Mouse genomic variation and its effect on phenotypes and gene regulation Nature 477 (2011) 289-294.

[3]   Y. Wu, E.G. Williams, S. Dubuis, A. Mottis, V. Jovaisaite, S.M. Houten, C.A. Argmann, P. Faridi, W. Wolski, Z. Kutalik, N. Zamboni, J. Auwerx, R. Aebersold, Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population Cell 158 (2014) 1415-1430.

[4]   J.M. Chick, S.C. Munger, P. Simecek, E.L. Huttlin, K. Choi, D.M. Gatti, N. Raghupathy, K.L. Svenson, G.A. Churchill, S.P. Gygi, Defining the consequences of genetic variation on a proteome-wide scale Nature 534 (2016) 500-505.

[5]   I.A. Maksakova, M.T. Romanish, L. Gagnier, C.A. Dunn, L.N. van de Lagemaat, D.L. Mager, Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line PLoS Genet 2 (2006) e2.

[6]   C. Stocking, C.A. Kozak, Murine endogenous retroviruses Cellular and molecular life sciences : CMLS 65 (2008) 3383-3398.

[7]   E.G. Williams, Y. Wu, P. Jha, S. Dubuis, P. Blattmann, C.A. Argmann, S.M. Houten, T. Amariuta, W. Wolski, N. Zamboni, R. Aebersold, J. Auwerx, Systems proteomics of liver mitochondria function Science 352 (2016) aad0189.

[8]   C. Qi, Y. Zhu, J. Pan, N. Usuda, N. Maeda, A.V. Yeldandi, M.S. Rao, T. Hashimoto, J.K. Reddy, Absence of spontaneous peroxisome proliferation in enoyl-CoA Hydratase/L-3-hydroxyacyl-CoA dehydrogenase-deficient mouse liver. Further support for the role of fatty acyl CoA oxidase in PPARalpha ligand metabolism J. Biol. Chem. 274 (1999) 15775-15780.

[9]   A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner Bioinformatics 29 (2013) 15-21.

[10]   J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer Nature biotechnology 29 (2011) 24-26.

[11]   H. Thorvaldsdottir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration Brief Bioinform 14 (2013) 178-192.

[12]   A. Le, A. Ng, T. Kwan, K. Cusmano-Ozog, T.M. Cowan, A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS) J Chromatogr B Analyt Technol Biomed Life Sci 944 (2014) 166-174.

[13]   M.E. Hinsdale, C.L. Kelly, P.A. Wood, Null allele at Bcd-1 locus in BALB/cByJ mice is due to a deletion in the short-chain acyl-CoA dehydrogenase gene and results in missplicing of mRNA Genomics 16 (1993) 605-611.

[14]   P.A. Andreux, E.G. Williams, H. Koutnikova, R.H. Houtkooper, M.F. Champy, H. Henry, K. Schoonjans, R.W. Williams, J. Auwerx, Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits Cell 150 (2012) 1287-1299.

[15]   M. van Weeghel, R. Ofman, C.A. Argmann, J.P. Ruiter, N. Claessen, S.V. Oussoren, R.J. Wanders, J. Aten, S.M. Houten, Identification and characterization of Eci3, a murine kidney-specific Delta3,Delta2-enoyl-CoA isomerase FASEB J 28 (2014) 1365-1374.

[16]   J.B. Cazier, P.J. Kaisaki, K. Argoud, B.J. Blaise, K. Veselkov, T.M. Ebbels, T. Tsang, Y. Wang, M.T. Bihoreau, S.C. Mitchell, E.C. Holmes, J.C. Lindon, J. Scott, J.K. Nicholson, M.E. Dumas, D. Gauguier, Untargeted metabolome quantitative trait locus mapping associates variation in urine glycerate to mutant glycerate kinase Journal of proteome research 11 (2012) 631-642.

[17]   Q. Qiao, T. Li, J. Sun, X. Liu, J. Ren, J. Fei, Metabolomic analysis of normal (C57BL/6J, 129S1/SvImJ) mice by gas chromatography-mass spectrometry: detection of strain and gender differences Talanta 85 (2011) 718-724.

[18]   M.G. Anderson, R.S. Smith, N.L. Hawes, A. Zabaleta, B. Chang, J.L. Wiggs, S.W. John, Mutations in genes encoding melanosomal proteins cause pigmentary glaucoma in DBA/2J mice Nat Genet 30 (2002) 81-85.

[19]   A.M. Ingalls, M.M. Dickie, G.D. Snell, Obese, a new mutation in the house mouse J Hered 41 (1950) 317-318.

[20]   Y. Zhang, R. Proenca, M. Maffei, M. Barone, L. Leopold, J.M. Friedman, Positional cloning of the mouse obese gene and its human homologue Nature 372 (1994) 425-432.

[21]   T.T. Huang, M. Naeemuddin, S. Elchuri, M. Yamaguchi, H.M. Kozy, E.J. Carlson, C.J. Epstein, Genetic modifiers of the phenotype of mice deficient in mitochondrial superoxide dismutase Hum Mol Genet 15 (2006) 1187-1194.

[22]   J. Nezu, I. Tamai, A. Oku, R. Ohashi, H. Yabuuchi, N. Hashimoto, H. Nikaido, Y. Sai, A. Koizumi, Y. Shoji, G. Takada, T. Matsuishi, M. Yoshino, H. Kato, T. Ohura, G. Tsujimoto, J. Hayakawa, M. Shimane, A. Tsuji, Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter Nat. Genet. 21 (1999) 91-94.

[23]   T. Yokoyama, D.W. Silversides, K.G. Waymire, B.S. Kwon, T. Takeuchi, P.A. Overbeek, Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice Nucleic Acids Res 18 (1990) 7293-7298.

Supplemental Figure legends

Figure S1. Identification of the structural variant in *Dhtkd1* causing 2-aminoadipic acidemia in C57BL/6J mice. (A) Distribution of the structural variant of interest in different inbred mouse strains. These data were obtained from the Mouse Genomes Project [1, 2]. D denotes deletion. (B) Genotyping by PCR amplification of the structural variant in a B6129PF2 cohort. (C) Quantitative visualization of splice junctions from RNA sequencing reads (Sashimi plot) mapping to *Dhtkd1*. RNA sequencing data are from livers of the B6129PF2 cohort. Mice are ordered according to their *Dhtkd1* genotype, i.e. B6/B6 (in blue), B6/129 (in red) or 129/129 (in green). Normalized library size for each sample is indicated between parentheses.

Figure S2. Quantification of the functional impact of the *Dhtkd1*[B6] variant using the B6129PF2 cohort. (A) RNA sequencing counts mapping to intron 4 of *Dhtkd1*, to all *Dhtkd1* exons and to the *Dhtkd1* splice junction between exon 4 and 5. (B) Quantitative PCR of total and functional *Dhtkd1* mRNA. Expression values were normalized using *Rplp0* or expressed as a ratio. All data were analyzed using linear regression analysis.

Figure S3. Identification of the structural variant in *Bckdhb* causing BCKDHB deficiency in C57BL/6J mice. (A) Sashimi plot of RNA sequencing reads mapping to *Bckdhb* in liver of one mouse from the B6129PF2 cohort with a *Bckdhb*[B6/B6] genotype. (B) Quantification of reads covering rs13480313 in *Bckdhb*[B6/129] mice. Data are displayed as a ratio of the reference T allele over the variant A allele. (C) RNA sequencing counts mapping to all *Bckdhb* exons. (D) Quantification of RNA sequencing reads joining alternative (alt) exon 1 and exon 2 of *Bckdhb*. (E) Quantification of RNA sequencing reads joining exon 1 and exon 2 of *Bckdhb*. (F) Quantification of RNA sequencing reads joining exon 1 and exon 4 of *Bckdhb*. The effect size of each *Bckdhb*[B6] allele was estimated using linear regression. (G) BCKDHB immunoblots used for the quantification of BCKDHB protein expression in Figure 2C. CS denotes citrate synthase.

Figure S4. Isovaleric acidemia in 129 mice. (A) Sashimi plot of RNA sequencing reads mapping to *Ivd* in liver of mice from the B6129PF2 cohort. One sample is displayed for each *Ivd*
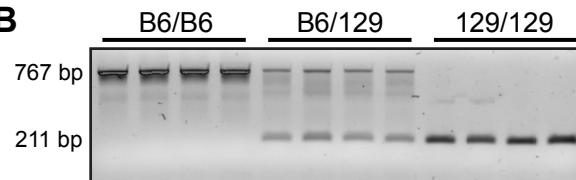
12

genotype; B6/B6 in red, B6/129 in blue and 129/129 in green. Normalized library size for each sample is indicated between parentheses. (B) Quantification of RNA sequencing reads joining exon 9 and 11 of *Ivd*. (C) RNA sequencing counts mapping to all *Ivd* exons. (D) Quantification of reads covering rs27440099 in *Ivd*$^{B6/129}$ mice. Data are displayed as a ratio of the variant A allele over the reference G allele. (E) Amplification of *Ivd* cDNA using a forward primer in exon 9 and a reverse primer in exon 11 shows skipping of exon 10 associated with the *Ivd*^{129} locus.

Figure S1

A

Chr2:5,924,595-5,925,150	
129P2/OlaHsd	D
129S1/SvImJ	D
129S5/SvEvBrd	D
A/J	
AKR/J	
BALB/cJ	D
C3H/HeJ	D
C57BL/6NJ	
CAST/EiJ	D
CBA/J	D
DBA/2J	D
FVB/NJ	
LP/J	D
NOD/ShiLtJ	
NZO/HILtJ	
PWK/PhJ	D
SPRET/EiJ	D
WSB/EiJ	D

B

C
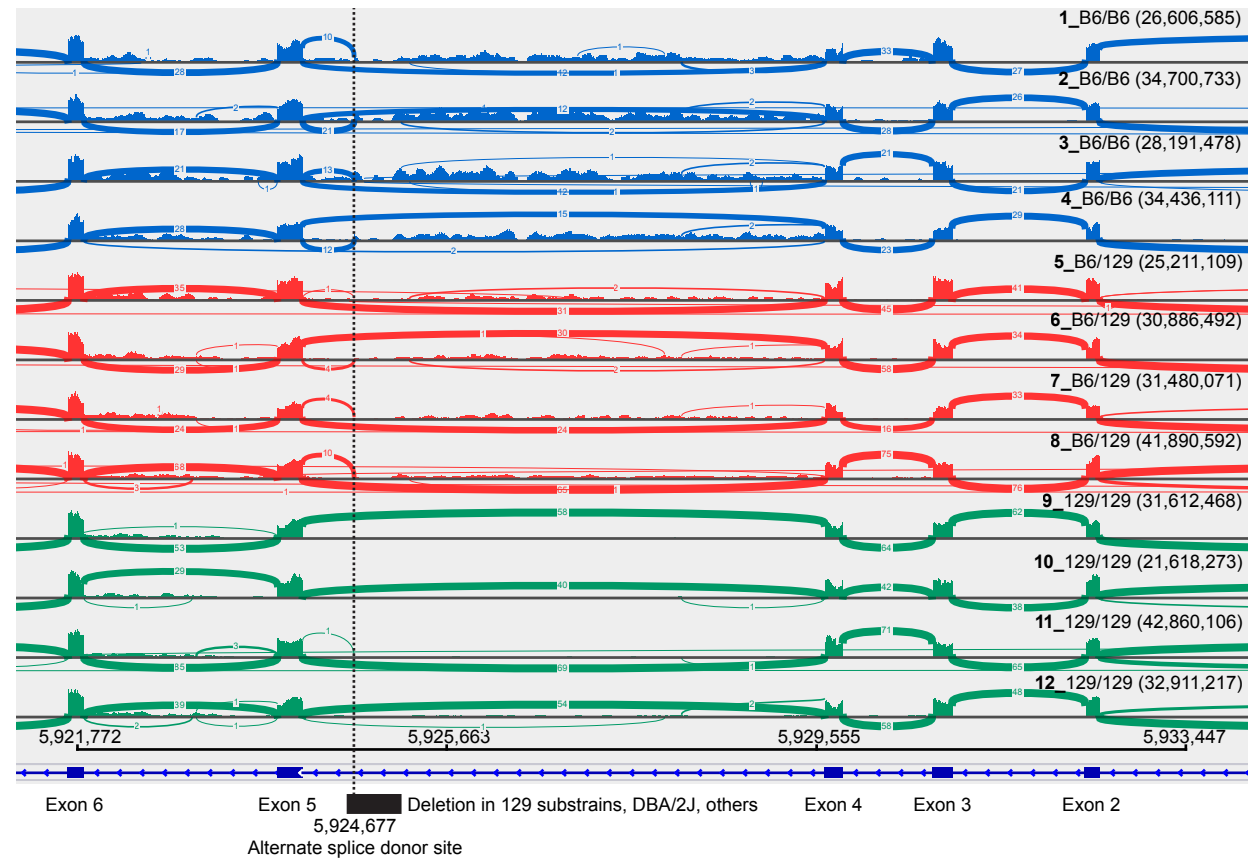
Figure S2

A RNA sequencing

intron 4

reads per million

Y = -6.748*X + 14.95
R square 0.9355
P < 0.0001

B6/B6 B6/129 129/129

Total exonic *Dhtkd1*

reads per million

Y = 20.76*X + 54.30
R square 0.7528
P = 0.0003

B6/B6 B6/129 129/129

Splice junction reads exon 4 to 5

reads per million

Y = 0.6597*X + 0.4327
R square 0.8946
P < 0.0001

B6/B6 B6/129 129/129

B quantitative PCR

Exon 5 (Total *Dhtkd1* mRNA)

2^(*Rplp0-Dhtkd1*)

Y = 0.230*X + 0.575
R square 0.5923
P = 0.0034

B6/B6 B6/129 129/129

Exon 4-5 (Functional *Dhtkd1* mRNA)

2^(*Rplp0-Dhtkd1*)

Y = 0.304*X + 0.419
R square 0.788
P = 0.0001

B6/B6 B6/129 129/129

Functional *Dhtkd1* / Total *Dhtkd1*

ratio

Y = 0.135*X + 0.735
R square 0.7915
P = 0.0001

B6/B6 B6/129 129/129

Figure S3

Figure S4

A

B

Joining 9-11

$Y = 1.730*X - 0.060$
R square 0.906
$P < 0.0001$

C

Total lvd exonic

D

rs27440099

$lvd^{B6/129}$

E

199bp
94bp