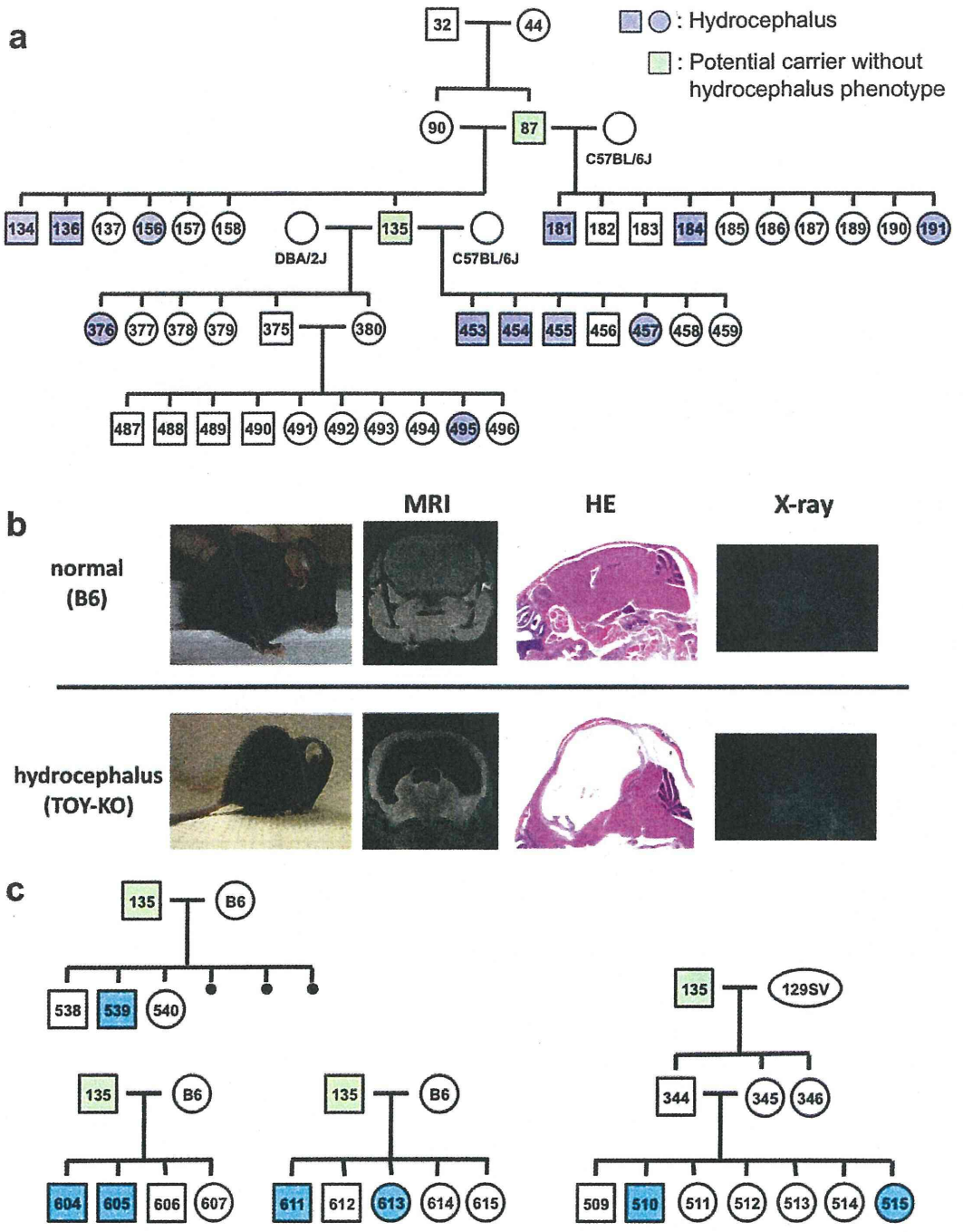

Figure 2 | Phenotypic variations observed in the progeny of TOY-KO mice. (a) The hydrocephalus trait was transmitted to the next generation in the TOY-KO pedigree. A hematoxylin/eosin-stained section showing the typical features of the hydrocephalus trait. Blue indicates a mouse with hydrocephalus, and green indicates a mouse carrying the causative mutation without the hydrocephalus phenotype (also shown in Supplementary Fig. S2 online). (b) Hydrocephalus. MRI, hematoxylin/eosin staining and X-ray images of normal (C57BL/6J) and hydrocephalus TOY-KO mice are shown in the upper panel. MRI images were obtained using an MRI mini SA (DS Pharma Biomedical Co. Ltd., Suita, Japan). X-ray images were obtained using a μFX-1000 (Fuji Photo File Co. Ltd.). (c) Pedigrees of the TOY-KO mouse mated with C57BL/6J (shown as B6) and 129Sv mice are shown in the lower panel. Blue indicates a mouse with hydrocephalus, and green indicates a mouse carrying the causative mutation without the hydrocephalus phenotype.

molecule for spontaneous G to T mutation in the mouse germ cell lineage. These mutations arose in all progeny of each generation and in all chromosomes that we analyzed (Figs. 4 and 5a). The mutations ranged from synonymous substitutions to harmful mutations, such as a gain of a stop codon in the *Ttn* gene responsible for human hypertrophic cardiomyopathy[18] (Supplementary Data S1 online).

By analyzing the position of the mutated G in di- and tri-nucleotide sequences, we found that G to T mutations occurred more often at GpC sites than at CpG sites, and tended to occur at tri-nucleotides, which are typical sequences found in triplet repeat expansion disorders (Fig. 5b, c), such as CAG (Huntington's disease), CTG (Myotonic dystrophy) and GAA (Friedreich ataxia)[19]. It is probable

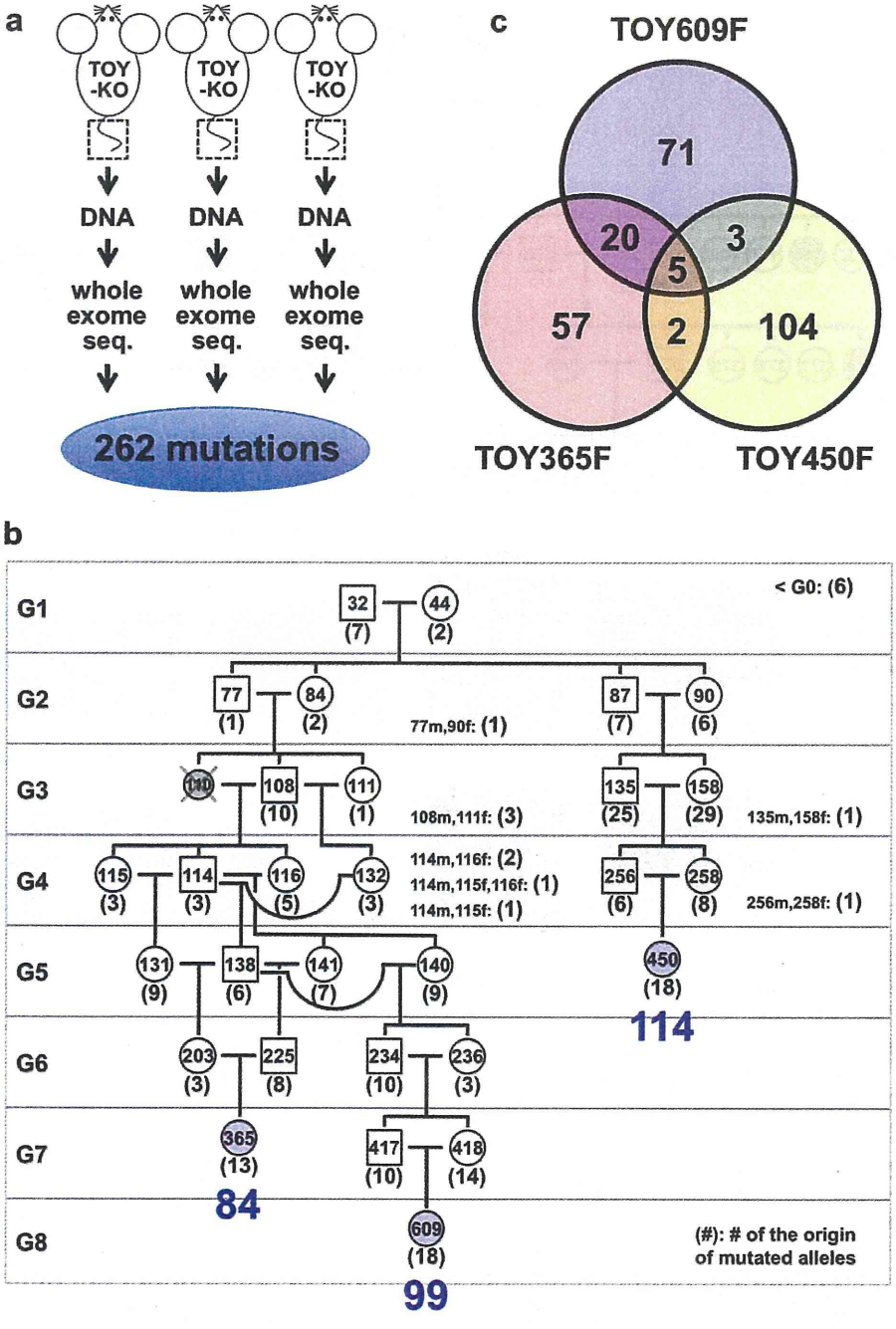

Figure 3 | Identification of *de novo* germline mutations in TOY-KO mice. (a) Scheme for screening of germline mutations. (b) Pedigree of TOY-KO mice used for germline mutation analysis. TOY365F, TOY609F and TOY450F were used to identify *de novo* germline mutations. Blue numbers, 84, 98, and 114, indicate the number of mutations detected in TOY365F, TOY609F and TOY450F, respectively. Numbers in parentheses indicate the number of original mutations in each generation, which were found in tail DNA for the first time in the pedigree. The DNA of TOY110F was unavailable; therefore, the mouse was excluded from the analysis. (c) The numbers of base substitution mutations found in TOY365F, TOY609F and TOY450F.

that uneven distribution of mutable 8-oxoG is reflected by the tendency for DNA oxidation, or by the site preference of DNA polymerases in incorporating 8-oxodGTP. We also detected two G to A and one A to G transition mutations that were classified as synonymous coding or intronic mutations (Table 1, Supplementary Data S1 online).

***De novo* germline mutation rate of TOY-KO mouse.** The detected mutations accumulated in TOY365F, TOY450F and TOY609F contained parts of the mutations that had occurred in the germ cells of the ancestral mice, because only half of the chromosomes derived from the father and mother had transmitted to the offspring via gametogenesis and fertilization in each generation. The numbers of newly arisen mutations detected only in TOY365F, TOY450F and TOY609F were 13, 18 and 18, respectively (Fig. 3b). Therefore, the *de novo* germline mutation rate was calculated to be 2.0×10^{-7}/base/generation ($13 + 18 + 18/3/40.9$ Mb $\times$ 2/generation). This mutation rate is 18-fold higher than the basal level, 1.1 $\times$

Table 1 | Spectrum of heritable mutations in TOY-KO mice

	All	G2–G8
G:C to A:T	7	2
A:T to G:C	2	1
G:C to T:A	252	244
A:T to C:G	1	0
G:C to C:G	0	0
A:T to T:A	0	0
total	262	247

Mutations detected in the 40.9 Mb exome sequences of TOY365F, TOY450F, and TOY609F (Fig. 3a) were classified into mutation types. The mutations observed in G2–G8 mice (Fig. 3b) were considered as mutations that occurred in the TOY-KO germ cell lineage.

10^{-8} mutations/base/generation, calculated from the specific locus test in mice[20]. For human trio analysis[1], the germline mutation rate was calculated to be 1.2×10^{-8} mutation/base/generation, and the G to T transversion mutation was observed in about 9% of all mutations. These results indicated that an approximately 200-fold increase in G to T transversion mutations occurred in the TOY-KO mice. No G to A transition mutations occurred in TOY365F, TOY450F, and TOY609F (totaling 245.4 Mb); therefore, the background mutation level of the TOY-KO mouse was estimated to be less than 4.1×10^{-9} G to A transition mutation/base/generation. This background mutation level is not high compared with that in humans (4.9×10^{-9} G to A transition mutation/base/generation)[1].

Fates of *de novo* germline mutations. By following up the mutated alleles in the pedigree, we observed the fates of the *de novo* mutations, in which some were fixed and others were eliminated in later generations. As shown in Fig. 6, for example, mutation #187 initially appeared in TOY108M (G3) as a heterozygous allele, indicating that the mutation probably occurred in the germ cell lineage of the parents, either TOY77M or TOY84F, and was transmitted to the progeny. At G5, it became homozygous in TOY138M and TOY-131F, and thus fixed in the progeny. Conversely, in another branch, the mutant allele was not transmitted to the offspring and eventually disappeared. These behaviors of the mutated allele represent the appearance, transmission, fixation and disappearance of a spontaneous mutation, which are the typical fates of a novel mutation in the evolutionary process.

Discussion

Little research has been performed to identify the causative molecule of spontaneous germline mutations because it is a rare event. We considered that the causative molecule must possess certain features that make DNA more prone to mutation, be generated endogenously and spontaneously and remain in the germ cell lineage. In 2006, we reported that endogenous 8-oxoG is distributed in the genome of human lymphocytes in the steady state[5]. We hypothesized that 8-oxoG also exists in the genome of germ lineage cells, and is responsible for spontaneous *de novo* germline mutations, because 8-oxoG is endogenously generated by ROS derived from cellular respiration, and is known to cause transversion mutations. By disruption of the

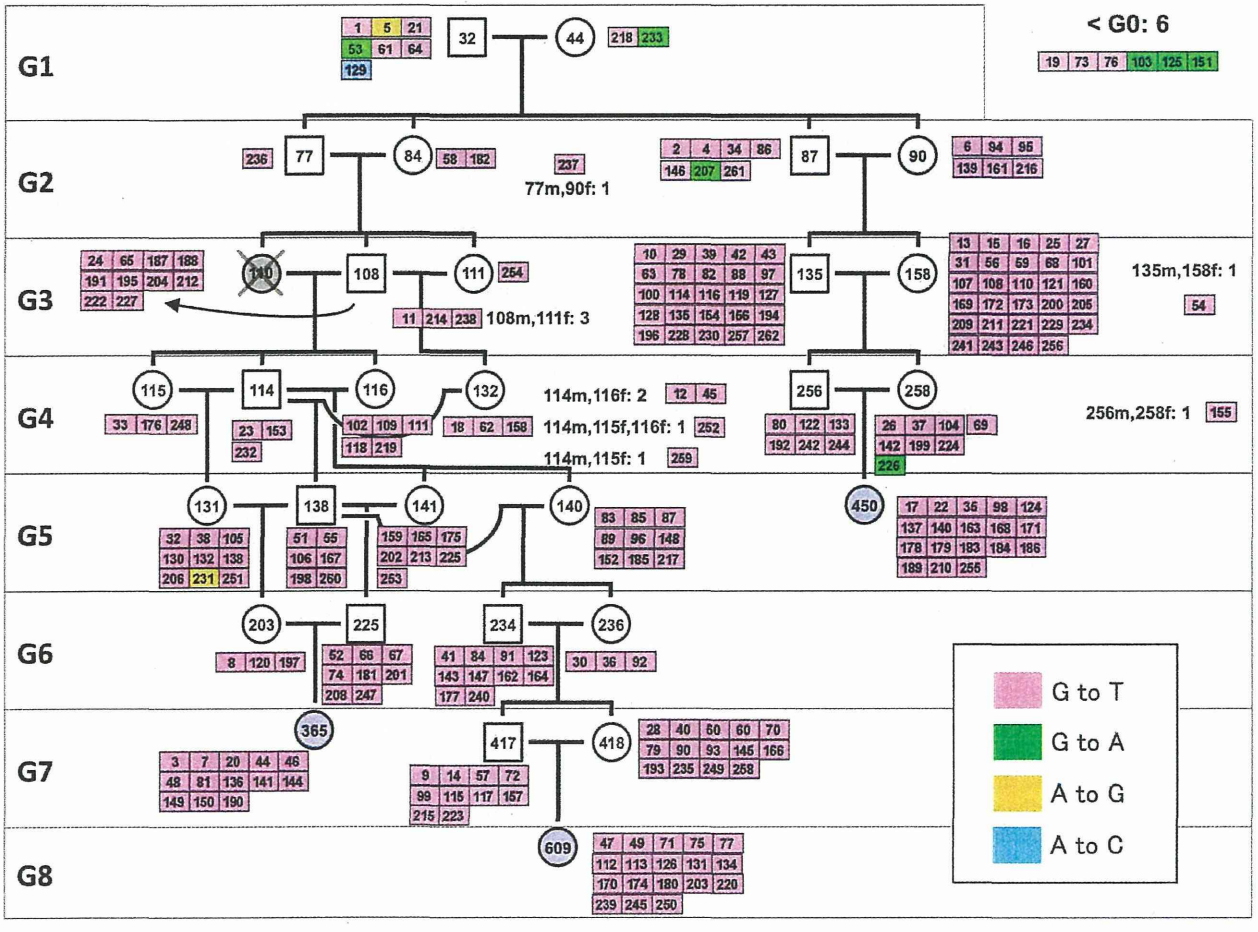

Figure 4 | Heritable mutations mapped in the pedigree of TOY-KO mice. The number in each box indicates the mutation ID number shown in Supplementary Data S1 online, and the color indicates the mutation category.

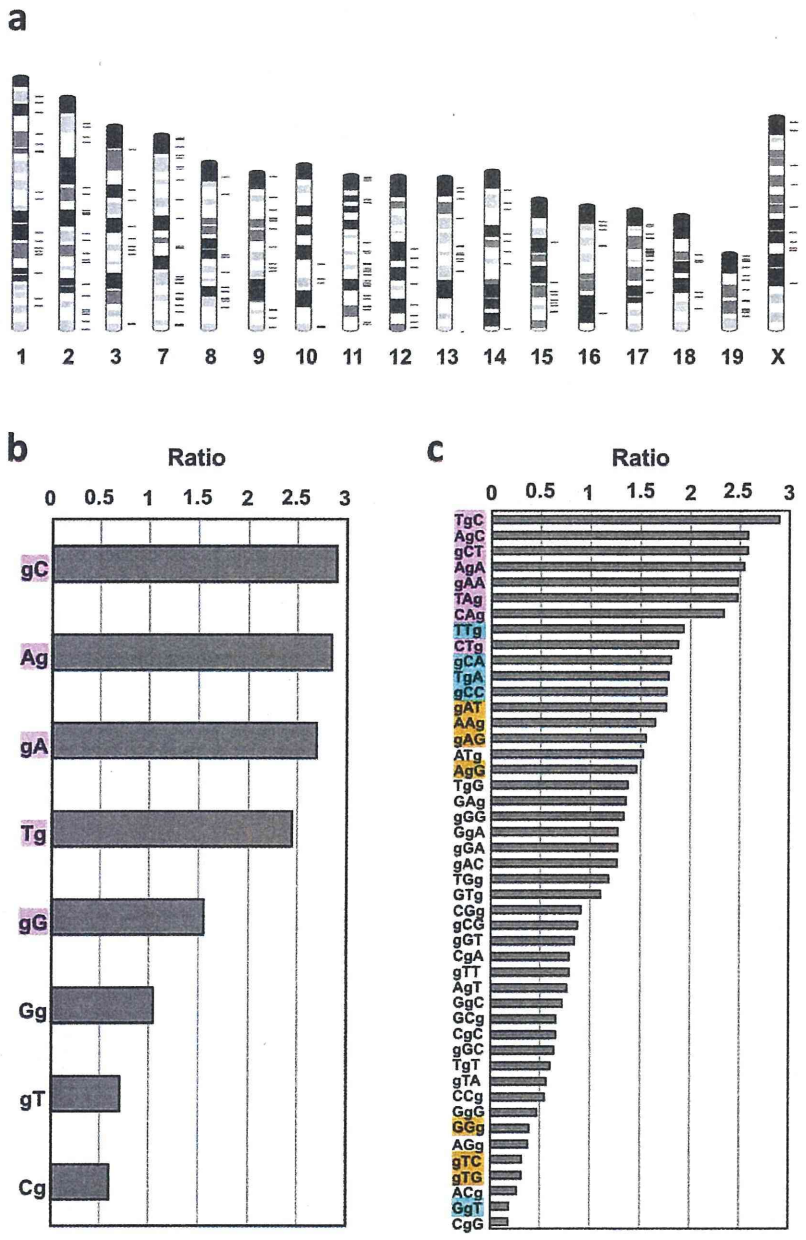

Figure 5 | Genome-wide distribution of mutations and site preferences of G to T mutations in di- and trinucleotide sequences. (a) Mutations detected in G2–G8 were mapped on a mouse G-band ideogram using Ideographica (http://www.ncrna.org/idiographica/). Each black transverse line on the right side of the chromosome represents a mutation site. (b) Site preferences of G to T mutations in di-nucleotide sequences. The plots represent the relative ratio of the actual value of detected mutations (G to T mutations in G2–G8) in each di-nucleotide to its occurrence level in the analyzed exome sequences. 'g' indicates the position of a mutated guanine. (c) Site preferences of G to T mutations in tri-nucleotides. For each nucleotide sequence, a chi square test (detected vs. expected) was performed, and the colored sequences indicate a significant difference: $P < 0.001$ (pink), $P < 0.01$ (blue), and $P < 0.05$ (orange).

8-oxoG exclusion system in mice, we detected increased spontaneous accumulation of germline mutations during the generations. These mutations were distributed throughout the chromosomes and inheritable to offspring across the generations, leading to an expansion of genetic diversity as well as disease-associated mutations.

The effects of 8-oxoG on spontaneous germline mutations were apparent in the TOY-KO mice. However, the production of 8-oxoG is dependent on the oxidation of guanine nucleotides, which occurs even in the wild-type cells independently of MTH1, OGG1 and MUTYH activities. It is likely that 8-oxoG universally causes *de novo* G-T transversion mutations, including germline mutations, although most of these mutations are efficiently prevented by the MTH1, OGG1 and MUTYH enzyme system.

When did the germline mutations occur? It is difficult to determine the timing of the occurrence of a mutation in the germ cell lineage; however, some examples were obtained that allowed us to speculate on the timing of mutations in our experiment. *De novo* mutations occur either in the germ cell lineage of the previous generation or during the very early developmental stage of the mutant mouse (Fig. 7). In eleven cases among 247 mutations, the mutations

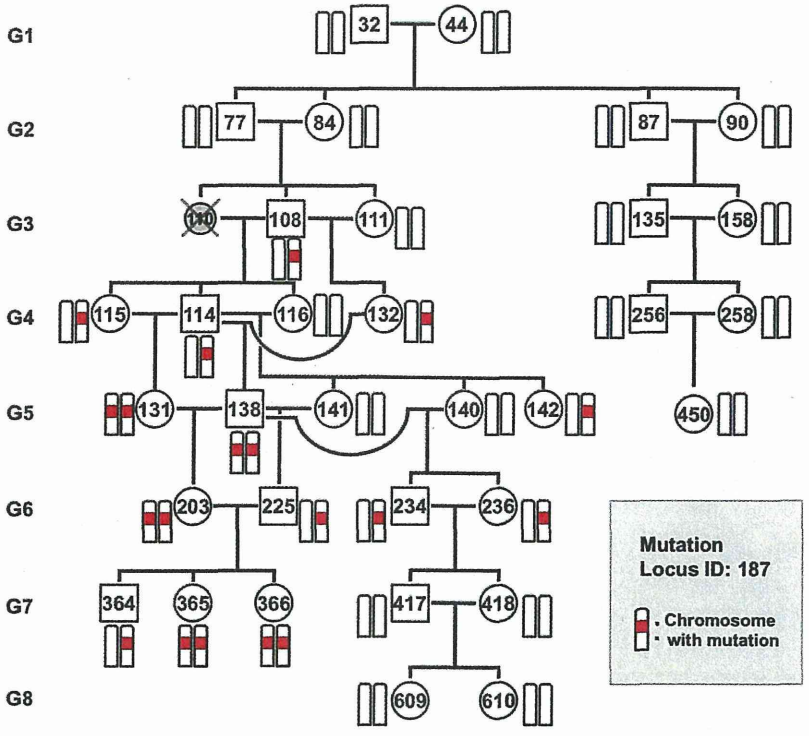

Figure 6 | Fate of a germline mutation. Mutation #187 (Ch. 15) was chosen to show the fate of a mutation generated in TOY-KO mice through the generations. This mutation initially appeared in TOY108M (G3) as a heterozygous allele. It was transmitted to progeny TOY-114M, TOY-115F, and TOY-132F. At G5, mutation #187 became homozygous in TOY138M and TOY131F, and thus were fixed in the progeny. Conversely, in another branch, the mutation was not transmitted from TOY-234M and TOY-236F (G6) to their offspring and eventually disappeared. The mutated locus is indicated in red.

had likely occurred in the germ cell lineage of the parents, because the original mutated allele was detected in multiple mice of the same generation (Fig. 3b). For three mutations on the X chromosome (Mutation ID #257, #261 and #262), which began in males with a heterozygous status (Supplementary Fig. S4 online), the mutation probably occurred in a cell at an early stage of embryonic development, resulting in mosaicism of tail tissue. These results showed that the germline mutations occurred at different developmental stages of the germ cell lineage. It is noteworthy that most germline mutations occurred during mitoses, because the germ cell lineage from fertilized egg to differentiated sperm or egg requires a large number of mitoses and only one meiosis. In the other cases (233/247) shown in Fig. 3b (G2–G8), the original mutated allele was found in a single mouse of each generation, and we could not identify when the mutation occurred.

By analogy to the *E. coli* system, we considered that 8-oxoG-induced G to T mutation is suppressed by OGG1, MUTYH, and MTH1, whereas the A to C mutation is prevented by MTH1 in mammalian cells (Supplementary Fig. S5 online). However, in contrast to the *E. coli mutT, mutM, mutY* triple mutant, in which both G to T and A to C mutations increased[21], no A to C germline mutations were detected in the TOY-KO mouse. Thus, it is likely that different mechanisms, such as mismatch repair[22] or proof reading by DNA polymerase, may function to avoid A to C mutations caused by 8-oxodGTP in the TOY-KO mouse, even in the absence of MTH1. It has been reported that 2-hydroxy-deoxyadenosine (2-OHdA), an oxidized form of deoxyadenosine, is recognized as a substrate by the MUTYH protein and possesses premutagenic features[23,24]. 2-OHdATP, a triphosphate form of 2-OHdA, is a substrate of the MTH1 protein[25]. The MutY and MutT proteins of *Escherichia coli* cannot recognize 2-OHdA, in contrast to the mammalian enzymes[24,26]. At the present, we cannot evaluate the contribution of 2-OHdA to the increase of germline mutation observed in TOY-KO mice, because we have not yet confirmed the accumulation of 2-OHdA in the DNA. Thus, the significance of 2-OHdA for germline mutations remains to be elucidated.

The TOY-KO mouse strain spontaneously accumulates mutations in the homozygous status. For genome-wide screening of mutants, this mouse has unique features and has the potential to take a complementary role to ENU mutagenesis[27,28]. The mutation is specific for G to T transversions, and occurs spontaneously and continuously in both male and female germ lineage cells of TOY-KO mice. The mutation rate of TOY-KO mice (0.2 mutation/Mb/generation, on average, in male and female) is lower than ENU-treated male gametes (1 mutation/0.42–1.82 Mb for male mouse[27], 1 mutation/3.7 Mb in male rat[28]); however, the number of mutations carried by each TOY-KO mouse increased as the generations increased. Similar to ENU mutagenesis, phenotype-driven screening is available. Currently, the TOY-KO mouse is only available in the C57BL/6J genetic background; however, it would be a good system for large genome-wide screening of dominant mutations. Using such mutator mice with a well-controlled genetic background would permit the evaluation of the contribution of aging and the difference between spermatogenesis and oogenesis on the accumulation of germline mutations. This system also enables us to assess the genotoxic effects of chemical and environmental factors on mammalian germ lineage cells.

Although *de novo* germline mutations cause sporadic genetic diseases in humans, their occurrence is an important step for the evolution of species, as well as selection for survival. 8-oxoG, one of the causative molecules of these mutations, is endogenously produced by ROS generated from biological processes, such as oxygen respiration

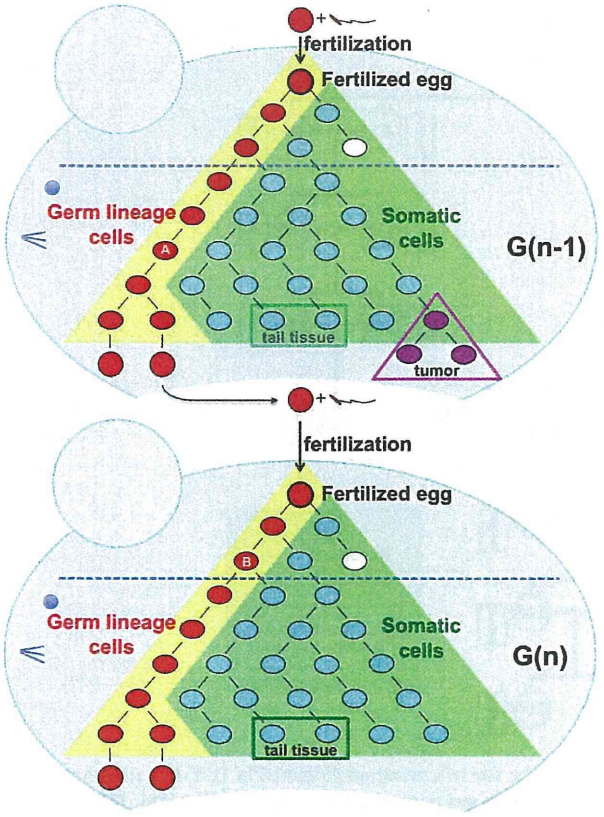

Figure 7 | Germ line mutations occur at different stages of the germ cell lineage. Mutations detected in the tail DNA of the first mutant mouse had occurred either in the germ lineage cells of the previous generation or during the very early developmental stage of the mutant mouse. Mutations start to accumulate from the first replication of fertilized egg DNA; however, each mutation is diluted out in the tissue DNA. Therefore, we used the tail DNA sequence as a reference sequence of fertilized egg DNA. In contrast to tail tissue, differentiated gametes can transmit their sequence information monoclonally to offspring. If the original mutated allele was mapped in multiple mice of the same generation, such as mutation #54 (in Fig. 4, Supplementary Data S1 online), the mutation probably occurred in the germ lineage cells of the parents (indicated as A). For mutations in the X chromosome (such as mutation #261), which began in the male with a heterozygous status (see Supplementary Fig. S4 online), the mutation probably occurred in a cell during the early stage of embryonic development (shown as B), resulting in mosaicism of tail tissue. These results indicate that germline mutations occur at different developmental stages of the germ cell lineage.

and inflammation, and is widely present in the DNA of various organisms. It is likely that the oxidative environment expands the genetic diversity of species by increasing the mutation rate of the germ lineage cells to accelerate the evolutionary process. MTH1, OGG1 and MUTYH, which are well conserved among species, may have contributed coordinately to control the germline mutation rate to an appropriate level for each species during evolution by controlling the amount of 8-oxoG in the genome (Supplementary Fig. S1 online).

Methods

Animals. *Mth1*$^{+/-}$, *Ogg1*$^{+/-}$, and *Mutyh*$^{+/-}$ mice were established[13,14,16] and backcrossed to C57BL/6J:Jcl (Clear Japan, Tokyo, Japan) for more than 16 generations. By crossing the C57BL/6J-background *Ogg1*$^{+/-}$, *Mth1*$^{+/-}$, and *Mutyh*$^{+/-}$ mice, we obtained *Mth1*$^{+/-}$/*Ogg1*$^{+/-}$ mice and *Ogg1*$^{+/-}$/*Mutyh*$^{+/-}$ mice. *Mth1*$^{+/-}$/*Ogg1*$^{+/-}$ mice were then mated with *Ogg1*$^{+/-}$/*Mutyh*$^{+/-}$ mice to obtain *Mth1*$^{+/-}$/*Ogg1*$^{-/-}$/*Mutyh*$^{+/-}$ mice. Finally, by crossing the *Mth1*$^{+/-}$/*Ogg1*$^{-/-}$/*Mutyh*$^{+/-}$ mice, we obtained a pair of *Mth1*$^{-/-}$/*Ogg1*$^{-/-}$/*Mutyh*$^{-/-}$ mice (TOY32M and TOY44F). All animals were maintained in a temperature-controlled (22 ± 2°C, 55 ± 5% humidity), specific pathogen-free room with a 12-h light-dark cycle. The care and use of all animals were performed in accordance with prescribed national guidelines, and the Animal Care and Use Committee of Kyushu University granted ethical approval for the study.

Quantification of 8-oxo-dG by LC-MS/MS. To detect the level of nuclear 8-oxodG, C57BL/6J:Jcl and TOY-KO mice (12–14 weeks old) were euthanized by cervical dislocation, and tissues were immediately removed and frozen in liquid nitrogen. DNA was extracted using a DNA Extractor TIS Kit (# 296-67701, Wako Pure Chemical Industries, Osaka, Japan), according to the manufacturer's instructions, with a slight modification: 10 mM 2, 2, 6, 6-tetramethylpiperidine-N-oxyl (Wako Pure Chemical Industries) was added to all reagents at all stages of manipulation[29]. Extracted DNA was hydrolyzed with 0.17 mg/ml nuclease P1 (Yamasa, Chiba, Japan) and 1.7 μM acid phosphatase (P-1435, Sigma-Aldrich Japan Inc., Tokyo, Japan) in 17 mM sodium acetate buffer (pH 4.5) at 37°C for 30 min, followed by filtration at 12,000 × *g* for 3 min (Ultrafree-MC probind 0.45 μm, Millipore, Billerica, MA). The digested samples (100 μl) were subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis using a Shimadzu VP-10 HPLC system connected to an API3000 MS/MS system (PE-SCIEX, SpectraLab Scientific Inc, Ontario, Canada).

Statistical analyses. Statistical analyses were conducted using JMP 9.02 (SAS Institute Japan, Tokyo, Japan).

Detection of germ line mutations by whole exome sequencing. Exome sequencing libraries for three TOY-KO mice (TOY365F, TOY450F and TOY609F) and five DBF1 (DBA/2J:Jcl × C57BL/6J:Jcl F1) mice as controls were prepared using a SureSelect^{XT} Mouse All Exon Kit (Agilent Technologies Japan, Tokyo, Japan), according to the manufacturer's instructions. Briefly, 3 μg of genomic tail DNA was sonicated into 150–180 bp fragments using a Covaris S2 System (Covaris, Woburn, MA, USA). The adaptors were ligated to the sonicated DNA after blunting and ~200 bp fragments were extracted using a 2% E-Gel (Life Technologies Japan, Tokyo, Japan). The extracted fragments were amplified with 2.5 mM SureSelect Pre-Capture primers and Platinum PCR Amplification Mix (Life Technologies), under the following conditions: 72°C for 20 min and 95°C for 5 min; 12 cycles of 95°C for 15 sec, 54°C for 45 sec and 70°C for 1 min; and a final extension at 70°C for 5 min. The PCR products were purified with a PureLink column (Life Technologies Japan). Purified PCR products (500 ng) were hybridized for 36 h at 65°C with SureSelect baits, according to the manufacturer's protocol. The captured libraries were amplified with the SureSelect Barcoding primer (BC1-8) for SOLiD with Herculase II Fusion DNA Polymerase (Agilent Technologies Japan), under the following conditions: 95°C for 5 min; 8 cycles of 95°C for 15 sec, 54°C for 45 sec and 70°C for 1 min; final extension at 70°C for 5 min. The captured barcoding libraries were quantified with an Agilent QPCR NHS Library Quantification Kit (Agilent Technologies Japan) and pooled. The four pooled libraries (1 pM) were amplified and purified with an EZ bead system (Life Technologies Japan). Purified P2-enriched beads were sequenced on one full slide of a SOLiD4 system (Life Technologies Japan). About 130 million paired-end sequencing reads (50 bp and 35 bp) were obtained from each library. Bioscope1.3.1 (Life Technologies Japan) was used to map the SOLiD paired-end reads to the mm9 reference mouse genome sequence (MGSCv37) using default parameters for Targeted resequencing methods. BEDtools v2.16.2 were used to calculate the coverage depth statistics and target enrichment efficiency. Avadis-NGS v1.3 (Strand Scientific Intelligence Inc., Karnataka, India) was used to carry out single nucleotide variant (SNV) calling with eight BAM format files (three TOY-KO lines and five control samples). The cutoff parameters of the SNV call were as follows: filtered sequencing quality ≤20, filtered PCR duplications, consensus base quality ≤50, total coverage <10, variants read depth <3, and the Decibel Score by Avadis-NGS v1.3 <50. The Decibel Score, read depth of the SNV allele and SNV allele frequency were used to sort these candidates. The iterative genomic viewer was used to check the candidates sequentially to eliminate apparent false positives. Finally, MassARRAY was used to select 286 mutation candidates for validation experiments (Supplementary Table S1 online).

Confirmation of mutations by sequencing. A MassARRAY3 Analyzer (Sequenom Inc, San Diego, CA) with iPLEX Gold Genotyping Reagent (Sequenom Inc) was used to validate the 286 candidates, according to the manufacturer's instructions. Briefly, MassARRAY Typer4 Assay Designer (Sequenom Inc) designed the 286 PCR primer pairs and 286 iPLEX primers as single-base extension primers for each candidate. We used 37 genomic DNA samples, including 35 samples from the TOY-KO pedigree and two control samples, as well as C57BL/6J and the original ES cell DNA to determine the origin of the *de novo* mutations in the TOY-KO pedigree. Ten nanograms of genomic DNA were used in each multiplex PCR for the MassARRAY. After dephosphorylation, single-base extension with the iPLEX primer and desalting were performed. The reaction products were spotted onto a 384-format SpectroCHIP with a MassARRAY Nanodispenser (Sequenom Inc) and then subjected to a MassARRAY 3 analyzer (Sequenom Inc). MassARRAY Typer 4.0 software (Sequenom Inc) was used to analyze the mass spectrum data.

ORIGINAL

Advance Publication
doi: 10.1507/endocrj. EJ14-0052

Association of polymorphisms in *GCKR* and *TRIB1* with nonalcoholic fatty liver disease and metabolic syndrome traits

Aya Kitamoto[1)], Takuya Kitamoto[1)], Takahiro Nakamura[2)], Yuji Ogawa[3)], Masato Yoneda[3)], Hideyuki Hyogo[4)], Hidenori Ochi[4)], Seiho Mizusawa[1)], Takato Ueno[5)], Kazuwa Nakao[6)], Akihiro Sekine[1)], Kazuaki Chayama[4)], Atsushi Nakajima[3)] and Kikuko Hotta[1)]

[1)] *Medical Research Support Center, Pharmacogenomics, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan*
[2)] *Laboratory for Mathematics, National Defense Medical College, Tokorozawa 359-8513, Japan*
[3)] *Division of Gastroenterology, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan*
[4)] *Department of Medicine and Molecular Science, Division of Frontier Medical Science, Programs for Biomedical Research, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima 734-8551, Japan*
[5)] *Research Center for Innovative Cancer Therapy, Kurume University, Kurume 830-0011, Japan*
[6)] *Medical Innovation Center, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan*

Abstract. In several genome-wide association studies, nonalcoholic fatty liver disease and alanine aminotransferase susceptibility variants have been identified in several genes, including *LYPLAL1*, *ZP4*, *GCKR*, *HSD17B13*, *PALLD*, *PPP1R3B*, *FDFT1*, *TRIB1*, *COL13A1*, *CPN1*, *ERLIN1*, *CWF19L1*, *EFCAB4B*, *PZP*, and *NCAN*. To investigate the relationship between these genes and nonalcoholic fatty liver disease in the Japanese population, we genotyped 540 patients and 1012 control subjects for 18 variations. We performed logistic regression analyses to characterize the association between the tested variations and nonalcoholic fatty liver disease. Metabolic syndrome and histological traits were also analyzed by linear regression. We also examined *GCKR* rs780094, *TRIB1* rs2954021, and *PNPLA3* rs738409 for epistatic effects. The A-allele of rs780094 in *GCKR* (P = 0.0024) and the A-allele of rs2954021 *TRIB1* ($P = 4.5\times10^{-5}$) were significantly associated with nonalcoholic fatty liver disease. *GCKR* rs780094 was also associated with decreased plasma glucose, and increased triglycerides in the patient and control groups. *GCKR* rs780094 was also associated with an increased ratio of visceral to subcutaneous fat area in the patients with nonalcoholic fatty liver disease. Variations in *GCKR*, *TRIB1*, and *PNPLA3* independently influenced nonalcoholic fatty liver disease and had no epistatic effects. Our data suggest variations in *GCKR* and *TRIB1* are involved in the development of nonalcoholic fatty liver disease.

Key words: *GCKR*, *TRIB1*, Nonalcoholic fatty liver disease, Metabolic syndrome, Japanese

NONALCOHOLIC FATTY LIVER DISEASE (NAFLD) has been recognized as an important health concern [1, 2]. NAFLD is a spectrum of liver diseases ranging from simple steatosis, through steatohepatitis (NASH), to fibrosis and ultimately cirrhosis. The frequency of patients presenting with NAFLD has increased in Japan in proportion to the increase in the population with metabolic syndrome [3]. NAFLD is observed in 20–30% of the population in Japan and approximately 1–3% of them are considered to have NASH, similar to American and European populations [3, 4].

In addition to environmental factors, genetic factors are important in the development of NAFLD [5]. In a previous search for these genetic factors, we found that variations in peroxisome proliferator-activated receptor γ coactivator 1α (*PPARGC1A*), angiotensin II type 1 receptor (*AGTR1*), and nitric oxide synthase 2 (inducible) (*NOS2*) are associated with NAFLD in Japanese individuals [6-8]. Genome-wide association studies (GWAS) have shown that SNPs in the patatin-like phospholipase domain containing 3 (*PNPLA3*) influence NAFLD and plasma liver enzymes [9-12]. We reported that the risk allele (G) of rs738409 in *PNPLA3* is strongly associated with NAFLD, as well as with

Submitted Feb. 3, 2014; Accepted Mar. 24, 2014 as EJ14-0052
Released online in J-STAGE as advance publication May 1, 2014
Correspondence to: Kikuko Hotta, M.D., Ph.D., Assistant Professor, Medical Research Support Center, Pharmacogenomics, Kyoto University Graduate School of Medicine, Yoshida-Konoecho, Sakyo-ku, Kyoto 606-8501, Japan.
E-mail: kikukoh@kuhp.kyoto-u.ac.jp

increases in aspartate aminotransferase (AST), alanine aminotransferase (ALT), and fibrosis stage in Japanese patients with NAFLD [13]. In a recent GWAS, we found that polymorphisms in the SAMM50 sorting and assembly machinery component (*SAMM50*), parvin, β (*PARVB*), and *PNPLA3* were associated with the development and progression of NAFLD [14].

Several other susceptibility loci for NAFLD [9-12, 14] and ALT [15-17] have been reported in GWAS, but these loci have not been confirmed in the Asian population. We investigated the association between SNPs identified by GWAS and NAFLD in the Japanese population.

Materials and Methods

Study subjects

The entire study was conducted in accordance with the guidelines of the Declaration of Helsinki. Written informed consent was obtained from each subject, and the protocol was approved by the ethics committee of Kyoto University, Yokohama City University, Hiroshima University, and Kurume University.

We enrolled 1012 control subjects (general population) from among Japanese volunteers undergoing medical examinations for common disease screening. Control subjects were retained from our previous study (control-2) [14]. Japanese patients with NAFLD who underwent liver biopsy (488 with NASH and 52 with simple steatosis) were enrolled; 392 (NAFLD-1) and 98 (NAFLD-2) of these subjects were retained from the previous study [14]. Control and NAFLD subjects were collected at Yokohama City University Hospital, Hiroshima University Hospital, and Kurume University Hospital. Patients with the following diseases were excluded from the study: viral hepatitis (hepatitis B and C, Epstein–Barr virus infection), autoimmune hepatitis, primary biliary cirrhosis, sclerosing cholangitis, hemochromatosis, α1-antitrypsin deficiency, Wilson's disease, drug-induced hepatitis, and alcoholic hepatitis (present or past daily consumption of more than 20 g alcohol per day). None of the patients showed clinical evidence of hepatic decompensation, such as hepatic encephalopathy, ascites, variceal bleeding, or a serum bilirubin level greater than two-fold the normal upper limit.

Liver biopsy tissues were stained with hematoxylin and eosin, reticulin, and Masson's trichrome stain. The histological criterion for NAFLD diagnosis is macrovesicular fatty change in hepatocytes with displacement of the nucleus toward the cell edge [18]. When more than 5% of hepatocytes are affected by macrovesicular steatosis, patients are diagnosed as having either steatosis or NASH; minimal criteria for the diagnosis of NASH included the presence of >5% macrovesicular steatosis, inflammation, and liver cell ballooning, typically with predominantly centrilobular (acinar zone 3) distribution [19, 20]. The degree of steatosis was graded as follows, based on the percentage of hepatocytes containing macrovesicular fat droplets: grade 0, no steatosis; grade 1, <33% hepatocytes containing macrovesicular fat droplets; grade 2, 33–66% of hepatocytes containing macrovesicular fat droplets; and grade 3, >66% of hepatocytes containing macrovesicular fat droplets [21]. Hepatitis activity (necroinflammatory grade) was also determined on the basis of the composite NAS, as described by Kleiner *et al.* [22]. NAS is the unweighted sum of the scores for steatosis, lobular inflammation, and hepatocellular ballooning, and ranges from 0 to 8. Fibrosis severity was scored according to the method of Brunt [18] and was expressed on a 4-point scale as follows: 0, none; 1, perivenular and/or perisinusoidal fibrosis in zone 3; 2, combined pericellular portal fibrosis; 3, septal/bridging fibrosis; and 4, cirrhosis.

Clinical and laboratory evaluation

Patient weight and height were measured using a calibrated scale after removing shoes and heavy clothing, if present. Venous blood samples were obtained after overnight fasting (12 h) to measure plasma glucose, hemoglobin A1c (HbA1c), total cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, serum AST, and ALT. All blood chemistry was measured using conventional methods.

The patients underwent CT imaging (in the supine position) to measure visceral fat area (VFA) and subcutaneous fat area (SFA) at the umbilical level (L4–L5); these values were calculated using the FatScan software program (N2system, Osaka, Japan) [23]. Clinical characteristics are shown in Table 1.

DNA extraction and SNP genotyping

Genomic DNA was extracted using Genomix (Talent Srl, Trieste, Italy) for blood samples collected from each subject. Invader probes (Third Wave Technologies, Madison, WI, USA) were designed for 18 SNPs previously identified as susceptibility loci for NAFLD [10, 11] or ALT [15-17]. The SNPs were

Table 1 Clinical characteristics

	NAFLD	Control	*P*
n	540	1012	–
Men/Women	285/255	500/512	0.21*
Type 2 diabetes (%)	250 (46.3%)	66 (6.5%)	1.1×10^{-76}*
Age (year)	50.5 ± 14.3	53.1 ± 15.3	0.0013
BMI (kg/m^2)	28.0 ± 5.0	22.7 ± 3.2	1.9×10^{-107}
Plasma glucose (mg/dL)	118.5 ± 36.1	98.2 ± 19.0	1.2×10^{-61}
HbA1c (%)	6.4 ± 1.3	5.5 ± 0.7	6.2×10^{-63}
Total cholesterol (mg/dL)	212.4 ± 39.3	208.5 ± 36.2	0.21
Triglycerides (mg/dL)	167.0 ± 105.2	110.0 ± 88.5	1.4×10^{-57}
HDL-cholesterol (mg/dL)	52.9 ± 14.8	62.7 ± 15.5	3.3×10^{-37}
SBP (mmHg)	128.1 ± 14.7	124.5 ± 19.1	1.0×10^{-4}
DBP (mmHg)	78.5 ± 11.6	76.3 ± 11.6	2.2×10^{-4}
AST (IU/L)	50.1 ± 29.9	23.0 ± 10.2	1.2×10^{-119}
ALT (IU/L)	81.0 ± 56.3	20.3 ± 11.8	4.1×10^{-171}
VFA (cm^2)†	128.6 ± 58.3	–	–
SFA (cm^2)†	209.9 ± 102.6	–	–
V/S ratio†	0.71 ± 0.39	–	–
Steatosis grade (1-3)	1.6 ± 0.7	–	–
Lobular inflammation (0-3)	1.3 ± 0.7	–	–
Hepatocyte ballooning (0-2)	1.1 ± 0.7	–	–
NAS (0-8)	4.0 ± 1.6	–	–
Fibrosis stage (0-4)	1.7 ± 1.0	–	–

P-values for comparison of the clinical data between the simple steatosis and NASH groups were obtained by the Mann–Whitney *U*-test. *, ratios were analyzed by χ^2-test. †, n = 439

as follows: rs12137855 near lysophospholipase-like 1 (*LYPLAL1*); rs2499604 near zona pellucida glycoprotein 4 (*ZP4*); rs780094 in glucokinase (hexokinase 4) regulator (*GCKR*); rs6834314 near hydroxysteroid (17-β) dehydrogenase 13 (*HSD17B13*); rs2710833 near palladin, cytoskeletal associated protein (*PALLD*); rs343062 in chromosome 7, position 35,549,066; rs4240624 and rs2126259 near protein phosphatase 1, regulatory subunit 3B (*PPP1R3B*); rs2645424 in farnesyl-diphosphate farnesyltransferase 1 (*FDFT1*); rs2954021 near tribbles homolog 1 (*TRIB1*); rs1227756 in collagen, type XIII, α 1 (*COL13A1*); rs10883437 and rs11597390 near carboxypeptidase N, polypeptide 1 (*CPN1*); rs2862954 in ER lipid raft associated 1 (*ERLIN1*); rs17668255 in CWF19-like 1, cell cycle control (*CWF19L1*); rs887304 in EF-hand calcium binding domain 4B (*EFCAB4B*); rs6487679 near pregnancy-zone protein (*PZP*); and rs2228603 in neurocan (*NCAN*). SNPs were genotyped by Invader assay as described [24], with a success rate of >98.0%.

Statistical analysis

We categorized the genotypes as 0, 1, or 2, depending on the number of risk alleles present. Odds ratios (OR) and *P*-values, adjusted for age, sex, logarithmically transformed body mass index (BMI), and the presence of type 2 diabetes mellitus (DM), were calculated by multiple logistic regression analysis. Multiple linear regression analyses were performed to test the independent effect of each allele on biochemical traits, and histological and anthropometric parameters, accounting for the effects of other variables (i.e., age, sex, and logarithmically transformed BMI). BMI, fasting plasma glucose, triglycerides, AST, ALT, VFA, SFA, and VFA to SFA (V/S) ratio values were logarithmically transformed before performing multiple linear and logistic regression analyses. Simple comparisons of the clinical data between NAFLD and control groups were carried out using the Mann–Whitney *U*-test. Male: female and the presence of DM ratios were analyzed by χ^2-test. To test SNP×SNP epistasis for case–control population-based samples, we used the logistic regression model for each SNP1 and SNP2, and fit the model in the form of $Y=\beta_0+\beta_1\times SNP1+\beta_2\times SNP2+\beta_3\times SNP1\times SNP2+\beta_4\times age+\beta_5\times sex+\beta_6\times \log_{10}(BMI)+\beta7\times DM$. Statistical analyses were performed using PLINK 1.07 (http://pngu.mgh.harvard.edu/purcell/plink) [25] and R software (http://www.r-project.

org/). *P*-values less than 2.8×10^{-3} (0.05/18) was considered significant.

Results

First, we examined the association of 18 SNPs with NAFLD and control subjects. We performed multiple logistic regression analysis using genotypes, age, sex, BMI, and the presence of DM as independent variables. Two SNPs, rs780094 in *GCKR* ($P = 0.0024$) and rs2954021 in *TRIB1* ($P = 4.5 \times 10^{-5}$), were significantly associated with NAFLD (Table 2). No other SNPs showed significant associations with NAFLD. Minor allele frequencies (MAFs) of eight SNPs were no more than 0.06. The lack of a significant association, especially in those SNPs with small MAF, is most likely due to the relatively low power of this study. All SNPs were in Hardy–Weinberg equilibrium ($P > 0.05$), with the exception of rs12137855 ($P = 0.025$) in the NAFLD patients.

Next, we examined the association of rs780094 and rs2954021 with metabolic syndrome traits in NAFLD. As reported previously [11], rs780094 in *GCKR* is associated with lower plasma glucose ($P = 0.0047$), higher triglycerides ($P = 0.0029$), and higher diastolic blood pressure ($P = 0.018$) (Table 3). An association

Table 2 Association tests of SNPs in patients with NAFLD and control subjects

Chr	SNP ID	position (build 132)	Nearby gene	Allele1/ allele2	Risk allele	Risk allele frequency		*P*-value	OR (95%CI)
						NAFLD	Control		
1	rs12137855	219,448,378	*LYPLAL1*	T/C	C	0.94	0.95	0.72	0.93 (0.61 - 1.40)
1	rs2499604	238,103,501	*ZP4*	A/G	A	0.46	0.48	0.59	0.95 (0.77 - 1.16)
2	rs780094	27,741,237	*GCKR*	G/A	A	0.62	0.55	**0.0024**	1.37 (1.12 - 1.68)
4	rs6834314	88,213,808	*HSD17B13*	G/A	A	0.69	0.66	0.30	1.12 (0.91 - 1.38)
4	rs2710833	169,409,958	*PALLD*	T/C	T	0.13	0.12	0.53	0.91 (0.67 - 1.23)
7	rs343062	35,549,066	no gene	T/C	T	0.46	0.47	0.92	0.99 (0.81 - 1.22)
8	rs4240624	9,184,231	*PPP1R3B*	G/A	A	0.99	0.99	0.40	0.64 (0.23 - 1.79)
8	rs2126259	9,185,146	*PPP1R3B*	A/G	G	0.99	0.99	0.37	0.62 (0.22 - 1.74)
8	rs2645424	11,684,463	*FDFT1*	C/T	C	0.23	0.21	0.30	1.13 (0.90 - 1.43)
8	rs2954021	126,482,077	*TRIB1*	A/G	A	0.52	0.45	**4.5×10^{-5}**	1.53 (1.25 - 1.88)
10	rs1227756	71,588,504	*COL13A1*	A/G	G	0.72	0.72	0.82	1.03 (0.82 - 1.29)
10	rs10883437	101,795,361	*CPN1*	A/T	T	0.83	0.82	0.64	0.94 (0.73 - 1.21)
10	rs11597390	101,861,435	*CPN1*	A/G	G	0.96	0.95	0.80	0.94 (0.59 - 1.51)
10	rs2862954	101,912,064	*ERLIN1*	C/T	T	0.97	0.95	0.58	1.16 (0.69 - 1.94)
10	rs17668255	102,000,701	*CWF19L1*	T/C	C	0.97	0.95	0.54	1.17 (0.70 - 1.97)
12	rs887304	3,757,548	*EFCAB4B*	A/G	A	0.005	0.001	0.70	1.46 (0.22 - 9.64)
12	rs6487679	9,371,332	*PZP*	C/T	C	0.11	0.11	0.75	1.05 (0.77 - 1.43)
19	rs2228603	19,329,924	*NCAN*	T/C	T	0.05	0.06	0.41	0.82 (0.51 - 1.32)

The OR for each SNP was adjusted simultaneously for age, sex, logarithmically transformed BMI, and the presence of DM. Bold entries indicate *P*-value < 0.05.

Table 3 Tests of association between significant SNPs and metabolic traits in NAFLD

SNP ID	rs780094				rs2954021			
	NAFLD		Control		NAFLD		Control	
	β	*P*-value	β	*P*-value	β	*P*-value	β	*P*-value
Plasma glucose	-0.018	**0.0047**	-0.007	**0.014**	-0.003	0.67	-0.003	0.29
Total cholesterol	0.182	0.94	1.464	0.34	2.663	0.28	0.187	0.90
Triglycerides	0.037	**0.0029**	0.039	**2.5×10^{-5}**	0.013	0.30	0.005	0.58
HDL-cholesterol	-0.588	0.49	-0.946	0.12	0.750	0.39	0.014	0.98
SBP	0.000	1.00	-0.280	0.69	-0.180	0.88	-1.102	0.13
DBP	2.222	**0.018**	-0.082	0.86	-1.115	0.24	-0.311	0.50
AST	0.005	0.72	0.009	0.21	-0.009	0.51	0.007	0.33
ALT	-0.008	0.58	0.019	**0.036**	-0.017	0.27	0.012	0.18

Data were derived from linear regression analysis. Values of FPG, triglycerides, AST, and ALT were logarithmically transformed. Each metabolic phenotype was adjusted simultaneously for age, sex, and logarithmically transformed BMI. Bold entries indicate *P*-value < 0.05.