analysis. The bases substitutions were called using SAMtools (version 0.1.7) (Li et al. 2009), and the indels were called using both SAMtools and Pindel algorithms (Ye et al. 2009).

Variant filtering and somatic variant identification

To pick out the high-confident somatic variants, we applied the following rigorous filters and rules to the data set (Supplemental Fig. S15). The first filter applied is the "quality filter." Variants with a mapping quality of 20 or more, a *phred*-like consensus quality of 20 or more, a base call quality of more than 10, and a sequence coverage of 10× or more for both the cell line and matched normal sample were considered as high-quality variants. The setting for the filter conditions were optimized by comparing common SNPs detected by BWA (Li and Durbin. 2009) with those genotyped using Affymetrix Human SNP Array 6.0 (Affymetrix), ensuring a high concordance (99.84%) across two analyses (Supplemental Fig. S16).

The second filter applied, referred to as the "somatic filter," seeks to pick out the somatically acquired variants. All the high-quality variants produced from the above steps were passed through the "somatic filter," and only those meeting the threshold were considered as the somatic variants. The mutant allele (nonreference allele) ratio was calculated as follows:

$$\textit{Mutant allele ratio} = \frac{\text{Count of non-reference bases}}{\text{Count of total bases}} \times 100\%$$

The setting for the "somatic filter" is as described in Supplemental Figure S15; for the cell line sample, it is required that four or more reads supporting the mutant allele and the mutant allele ratio should be 15% or more. Moreover, the mutant allele should be supported by reads that aligned in both the forward and reverse directions. For the matched normal sample, given the potential sequencing errors and mapping errors, the mismatch should not be detected in more than 3% of the aligned reads and should not be detected in more than two reads. The indel, however, should not be detected in any of the aligned reads.

The third filter, referred to as the "false-positive filter," was then applied. This filter is used to remove the potential false-positive events that result from the homologous sequences within the human genome, mapping errors, and so on. For each of the somatic mutations produced in the above steps, we extracted 200–300 bases of DNA sequences flanking its mutation locus and mapped the sequences to hg18 using the BLAT algorithm. Subsequently, the mutations identified within the regions rich for homologous sequences were removed from the list. The somatic mutations were further examined using the integrated genome viewer (IGV), and any mutations found in a "noisy" background (multiple mismatches or indels in flanking sequences) were removed from the list.

As for detection of indels, one more step, called "rescue," was applied since the sequence read carrying a long indel toward its end is usually difficult to be aligned properly. We use the Pindel algorithm to rescue those possibly missed indels.

Variant annotation

Functional effects of filtered somatic variants were predicted using the SIFT algorithm (Kumar et al. 2009; http://sift.jcvi.org). The SIFT algorithm predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.

Mutation rate calculation and normalization

The background mutation rate (mutations/per Mb coding sequences) was calculated as follows:

$$\frac{\text{Sum of somatic mutations}}{\text{Sum length of exome targets} \times \text{number of tumor cell lines}}$$

The mutation rate of each gene was normalized by the frequency of mutations and the length of its coding sequences. Only somatic deleterious mutations, including missense substitutions, nonsense substitutions, frame-shift indels, and focal homozygous deletions were counted. The normalized mutation rate for each gene was calculated as follows, and a priority list was made accordingly:

$$\frac{\text{Sum of somatic mutations identified in the gene}}{\text{Sum length of coding regions of the gene} \times \text{number of tumor cell lines}}$$

Pathway analysis

The genes with somatic mutations were classified into different functional pathways using the Gene Ontology (GO) database (http://www.geneontology.org/). Only somatic deleterious mutations were counted. The normalized mutation rate for each pathway was calculated as below:

$$\frac{\text{Sum of somatic mutations identified in genes included}}{\text{Sum length of coding regions of genes included} \times \text{number of tumor cell lines}}$$

mRNA-seq and data analysis

Library preparation and mRNA-seq

Total RNA was extracted from PDAC-derived cell lines and the HPNE cells using the protocol outlined in the RNeasy Kit (Qiagen). Total RNA integrity was measured using a 2100 Bioanalyzer (Agilent Technologies), and all samples were confirmed to have an RNA Integrity Number (RIN) greater than 8.0 prior to further analysis. The mRNA-seq libraries were prepared using a paired-end mRNA Sequencing Sample Prep Kit (Illumina) following the manufacturer's protocols with slight modifications. Briefly, 2 μg of total RNA was used as the starting material, and the polyadenylated RNAs were selected using Sera-Mag Magnetic Oligo(dT) Beads (Illumina). The Poly(A)$^+$ RNA was then fragmented by heating for 90 sec at 94°C in the supplied fragmentation buffer. Fragmented RNA was mixed with random primers, incubated for 5 min at 65°C, and placed on ice briefly before starting cDNA synthesis. First-strand cDNA synthesis was performed using SuperScript II, and second-strand cDNA synthesis was performed using DNA Pol I in the supplied GEX second-strand reaction buffer. Subsequently, cDNA ends were repaired, and adenine was added to the 3′ end of the cDNA fragments to allow adaptor ligation. Paired-end adaptors were ligated to the cDNA fragments. The ligated product was run on a 2% agarose gel, and a 300 ± 20 bp fragment was cut out and extracted. PCR (eight cycles) was performed with Phusion High-Fidelity DNA Polymerase (Finnzymes Oy) following the manufacturer's protocols. The PCR products were cleaned up with Agencourt AMPure XP magnetic beads (Beckman Coulter); 6.0–6.7 pM/lane cDNA was applied to the flow cell and paired-end 76-nt-long reads were generated using Illumina GAIIx. Each sample was run on two lanes of Illumina flow cell.

Data alignment

All PF reads were aligned to hg18 using TopHat spliced aligner (Trapnell et al. 2009). Meanwhile, all PF reads were aligned to NCBI Reference Sequence (RefSeq) mRNA sequences using BWA. A merged file was generated for each sample by integrating the output of

TopHat with that of BWA for an optimal alignment for each sequence read. The ambiguously mapped reads and the duplicates were excluded. The level of gene expression was calculated in reads per kilobase of exonic sequence per million aligned reads (RPKM).

Mutation validation

For each of the somatic mutations identified by Exome-seq, we extracted the aligned mRNA-seq reads at its corresponding locus and examined if the mutant allele was also present in the cDNA sequences. The substitutions were called using SAMtools (Li et al. 2009). The small indels were called by both the SAMtools and Pindel algorithms (Ye et al. 2009). We focus on those loci covered by at least five reads, since it is rather difficult to call the variant accurately for poorly expressed genes. The mutation is supposed to be verified by mRNA-seq if at least two reads carried the mutant allele, and the mutant allele was detected in no less than 5% of the total reads aligned. For those loci covered by less than five reads but two or more reads, the mutation was also supposed to be verified if at least two reads carried the mutant allele.

Genome-wide SNP genotyping and DNA copy-number analysis

Genome-wide SNP genotyping was performed using the Affymetrix Genome-wide Human SNP Array 6.0 (Affymetrix) according to the manufacturer's instructions. SNPs were genotyped using the Birdseed version 2 module of the Affymetrix Genotyping Console software GTC 4.0.1, together with data from 45 HapMap-JPT samples (CEL files obtained from Affymetrix). DNA copy-number changes were analyzed using the Genome Imbalance Map (GIM) algorithm, as we previously described (Ishikawa et al. 2005).

The conventional MSI assay

The conventional MSI assay was performed using the proposed "Bethesda" panel of fluorescence-labeled markers, including *BAT25*, *BAT26*, *D2S123*, *D5S346*, and *D17S250* and an additional two markers, *NR21* and *NR27*. The primer sequences and PCR conditions have been previously described (Murayama-Hosokawa et al. 2010). In this study, we selected an additional three coding microsatellites and designed 6-carboxyfluorescein–labeled primers. Sequences of oligonucleotide primers for these three microsatellites are listed in Supplemental Table S4. PCR reactions were performed using the previously described reagents (Murayama-Hosokawa et al. 2010) under the following thermal cycle conditions: initial denature for 2 min at 94°C, followed by 32 cycles of denature for 15 sec at 94°C, annealing for 30 sec at 58°C, and primer extension for 30 sec at 68°C; the final extension step was carried out for 2 min at 68°C. After PCR, 1 μL of the properly diluted PCR product was mixed with 10 μL of Hi-Di Formamide and GeneScan 500 LIZ Size Standard (Applied Biosystems) mixture (37:1). This product was then denatured for 5 min at 95°C and put on ice immediately for 5 min before loading onto ABI 3130xl Genetic Analyzer (Applied Biosystems). The output data files were analyzed by GeneMapper Software Version 4.0 (Applied Biosystems). Determination of MSI status was made according to the presence of mutant alleles in tumor DNA compared with matched normal DNA.

MSI analysis by Exome-seq

We established a data analysis pipeline to identify small indels in the microsatellites. For each of the somatic indels identified in this study, we extracted the 50 bases of DNA sequences flanking its locus and examined if the indel was present in microsatellite sequences. Only those indels detected in the protein-coding microsatellites with at most 6 nt and repeated at least five times for mono- and dinucleotide microsatellites and at least three times for multiple-nucleotide microsatellites were counted. As shown in Figure 4, a graph was plotted for the indels in coding microsatellites according to the lengths of the indels and the number of sequence reads that supported the mutant alleles or the wild-type alleles. The microsatellite was suggested to be instable if a shorter allele (deletion) or a longer allele (insertion) was detected only in the tumor DNA. The sequence homology of each supporting read was further examined by the BLAT algorithm, and the reads rich of homologous sequences were discarded. The mutant allele ratio was then calculated using a formula as mentioned above.

MLH1 promoter methylation analysis

The methylation status of *MLH1* promoter was quantitatively measured using MassARRAY (Sequenom), as previously described (Yagi et al. 2010). Briefly, 500 ng gDNA was bisulfite converted using an EZ DNA Methylation Kit (Zymo Research) according to the manufacturer's instruction manual. Bisulfite-treated DNA was PCR amplified, and the PCR product was transcribed by in vitro transcription (IVT) prior to cleavage using RNase A. Unmethylated cytosine was converted to uracil by bisulfite treatment, while the methylated cytosine was not converted. Methylation status was then determined by the mass difference between A and G in the cleaved RNA product. Quantitative methylation scores were obtained at each analytic unit of a cleaved product, referred to as "CpG unit." The amplified DNA that was not methylated at all in any CpG sites was used as an unmethylated (0%) control. The amplified DNA, methylated by SssI methylase, was used as a fully methylated (100%) control.

Sanger sequencing

Oligo primers were designed to amplify the genome fragments containing the candidate nucleotide mutations from tumor cell line DNA and the matched normal DNA. PCR was performed using the high-fidelity DNA polymerase KOD-plus (TOYOBO) under optimized thermal conditions. PCR products were evaluated on a 2% agarose gel, purified and sequenced in both directions using Big Dye Terminator reactions, and subsequently loaded on an ABI 3130xl capillary sequencer (Applied Biosystems).

Statistical analysis

The *P*-value was calculated by Student's *t*-test when the data were normally distributed or by the nonparametric Wilcoxon signed-rank test when the data were not normally distributed. *P*-values less than 0.05 were considered to be statistically significant.

Data access

The Exome-seq data, mRNA-seq data, and SNP array data have been submitted to the European Genome-Phenome Archive (EGA; http://www.ebi.ac.uk/ega/), which is hosted at the European Bioinformatics Institute (EBI), under accession no. EGAS00001000149.

Acknowledgments

We thank Dr. Teruhiko Yoshida for having organized collaboration on the Exome-seq of the RCC project. We thank Ms. Kaori Shiina, Ms. Hiroko Meguro, Ms. Kaoru Nakano, and Ms. Saori Kawanabe for their excellent technical assistance. We acknowledge Dr. Michael Jones for the critical reading of the manuscript. This study was supported by Grants-in-Aid for Scientific Research (H.A.) and Scien-

tific Research on Priority Areas (H.A., T.N.); a grant for Translational Systems Biology and Medicine Initiative (TSBMI; H.A.) from the Ministry of Education, Culture, Sports, Science and Technology; the NFAT project from the New Energy and Industrial Technology Development Organization (NEDO; H.A., T.N.), Japan; and the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO; T.S., H.S.,Y.K.).

References

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, et al. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat* **30:** 757–770.

Biesecker LG, Shianna KV, Mullikin JC. 2011. Exome sequencing: the expert view. *Genome Biol* **12:** 128. doi: 10.1186/gb-2011-12-9-128.

Bodmer W, Bielas JH, Beckman RA. 2008. Genetic instability is not a requirement for tumor development. *Cancer Res* **68:** 3558–3560.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, et al. 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58:** 5248–5257.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, et al. 1994. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368:** 258–261.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci* **106:** 19096–19101.

Cirulli ET, Singh A, Shianna KV, Ge D, Smith JP, Maia JM, Heinzen EL, Goedert JJ, Goldstein DB. 2010. Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. *Genome Biol* **11:** R57. doi: 10.1186/gb-2010-11-5-r57.

Cunningham JM, Christensen ER, Tester DJ, Kim CY, Roche PC, Burgart LJ, Thibodeau SN. 1998. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res* **58:** 3455–3460.

Edelmann W, Cohen PE, Kane M, Lau K, Morrow B, Bennett S, Umar A, Kunkel T, Cattoretti G, Chaganti R, et al. 1996. Meiotic pachytene arrest in MLH1-deficient mice. *Cell* **85:** 1125–1134.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75:** 1027–1038.

Fujii K, Miyashita K, Yamada Y, Eguchi T, Taguchi K, Oda Y, Oda S, Yoshida MA, Tanaka M, Tsuneyoshi M. 2009. Simulation-based analyses reveal stable microsatellite sequences in human pancreatic cancer. *Cancer Genet Cytogenet* **189:** 5–14.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* **446:** 153–158.

Harada T, Chelala C, Bhakta V, Chaplin T, Caulee K, Baril P, Young BD, Lemoine NR. 2008. Genome-wide DNA copy number analysis in pancreatic cancer using high-density single nucleotide polymorphism arrays. *Oncogene* **27:** 1951–1960.

Hatch SB, Lightfoot HM Jr, Garwacki CP, Moore DT, Calvo BF, Woosley JT, Sciarrotta J, Funkhouser WK, Farber RA. 2005. Microsatellite instability testing in colorectal carcinoma: choice of markers affects sensitivity of detection of mismatch repair-deficient tumors. *Clin Cancer Res* **11:** 2180–2187.

Hemminki A, Peltomaki P, Mecklin JP, Jarvinen H, Salovaara R, Nystrom-Lahti M, de la Chapelle A, Aaltonen LA. 1994. Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet* **8:** 405–410.

Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. 2004. Nonsense-mediated decay approaches the clinic. *Nat Genet* **36:** 801–808.

Ishikawa S, Komura D, Tsuji S, Nishimura K, Yamamoto S, Panda B, Huang J, Fukayama M, Jones KW, Aburatani H. 2005. Allelic dosage analysis with genotyping microarrays. *Biochem Biophys Res Commun* **333:** 1309–1314.

Issaq SH, Lim KH, Counter CM. 2010. Sec5 and Exo84 foster oncogenic ras-mediated tumorigenesis. *Mol Cancer Res* **8:** 223–231.

Jiricny J. 1998. Eukaryotic mismatch repair: an update. *Mutat Res* **409:** 107–121.

Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321:** 1801–1806.

Kim MP, Fleming JB, Wang H, Abbruzzese JL, Choi W, Kopetz S, McConkey DJ, Evans DB, Gallick GE. 2011. ALDH activity selectively defines an enhanced tumor-initiating cell population relative to CD133 expression in human pancreatic adenocarcinoma. *PLoS ONE* **6:** e20636. doi: 10.1371/journal.pone.0020636.

Kolodner RD, Marsischky GT. 1999. Eukaryotic DNA mismatch repair. *Curr Opin Genet Dev* **9:** 89–96.

Kuismanen SA, Holmberg MT, Salovaara R, de la Chapelle A, Peltomaki P. 2000. Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. *Am J Pathol* **156:** 1773–1779.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4:** 1073–1081.

Lengauer C, Kinzler KW, Vogelstein B. 1998. Genetic instabilities in human cancers. *Nature* **396:** 643–649.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* **42:** 969–972.

Ma S, Chan YP, Woolcock B, Hu L, Wong KY, Ling MT, Bainbridge T, Webber D, Chan TH, Guan XY, et al. 2009. DNA fingerprinting tags novel altered chromosomal regions and identifies the involvement of SOX5 in the progression of prostate cancer. *Int J Cancer* **124:** 2323–2332.

Maitra A, Hruban RH. 2008. Pancreatic cancer. *Annu Rev Pathol* **3:** 157–188.

Marra G, Schar P. 1999. Recognition of DNA alterations by the mismatch repair system. *Biochem J* **338:** 1–13.

Metzker ML. 2010. Sequencing technologies—the next generation. *Nat Rev Genet* **11:** 31–46.

Murayama-Hosokawa S, Oda K, Nakagawa S, Ishikawa S, Yamamoto S, Shoji K, Ikeda Y, Uehara Y, Fukayama M, McCormick F, et al. 2010. Genome-wide single-nucleotide polymorphism arrays in endometrial carcinomas associate extensive chromosomal instability with poor prognosis and unveil frequent chromosomal imbalances involved in the PI3-kinase pathway. *Oncogene* **29:** 1897–1908.

Nakahori S, Yokosuka O, Ehata T, Chuang WL, Imazeki F, Ito Y, Ohto M. 1995. Detection of hepatitis B virus precore stop codon mutants by selective amplification method: frequent detection of precore mutants in hepatitis B e antigen positive healthy carriers. *J Gastroenterol Hepatol* **10:** 419–425.

Negrini S, Gorgoulis VG, Halazonetis TD. 2010. Genomic instability: an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11:** 220–228.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461:** 272–276.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, et al. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42:** 30–35.

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* **321:** 1807–1812.

Qiu W, Tong GX, Manolidis S, Close LG, Assaad AM, Su GH. 2008. Novel mutant-enriched sequencing identified high frequency of PIK3CA mutations in pharyngeal cancer. *Int J Cancer* **122:** 1189–1194.

Raschle M, Marra G, Nystrom-Lahti M, Schar P, Jiricny J. 1999. Identification of hMutLβ, a heterodimer of hMLH1 and hPMS1. *J Biol Chem* **274:** 32368–32375.

Schuster SC. 2008. Next-generation sequencing transforms today's biology. *Nat Methods* **5:** 16–18.

Shibata T, Saito S, Kokubu A, Suzuki T, Yamamoto M, Hirohashi S. 2010. Global downstream pathway analysis reveals a dependence of oncogenic NF-E2–related factor 2 mutation on the mTOR growth signaling pathway. *Cancer Res* **70:** 9095–9105.

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* **314:** 268–274.

Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* **458:** 719–724.

Sugarbaker DJ, Richards WG, Gordon GJ, Dong L, De Rienzo A, Maulik G, Glickman JN, Chirieac LR, Hartman ML, Taillon BE, et al. 2008.

Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc Natl Acad Sci* **105:** 3521–3526.
Suter CM, Martin DI, Ward RL. 2004. Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* **36:** 497–501.
Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, Garraway LA, LaFramboise T, Lee JC, Shah K, et al. 2006. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* **12:** 852–855.
Toma MI, Grosser M, Herr A, Aust DE, Meye A, Hoefling C, Fuessel S, Wuttig D, Wirth MP, Baretton GB. 2008. Loss of heterozygosity and copy number abnormality in clear cell renal cell carcinoma discovered by high-density affymetrix 10K single nucleotide polymorphism mapping array. *Neoplasia* **10:** 634–642.
Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, et al. 2011. High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* **43:** 464–469.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.
Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, et al. 2011. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* **469:** 539–542.
Vogelstein B, Kinzler KW. 2004. Cancer genes and the pathways they control. *Nat Med* **10:** 789–799.
Wilentz RE, Goggins M, Redston M, Marcus VA, Adsay NV, Sohn TA, Kadkol SS, Yeo CJ, Choti M, Zahurak M, et al. 2000. Genetic, immunohistochemical, and clinical features of medullary carcinoma of the pancreas: A newly described and characterized entity. *Am J Pathol* **156:** 1641–1651.
Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, et al. 2007. The genomic landscapes of human breast and colorectal cancers. *Science* **318:** 1108–1113.
Yagi K, Akagi K, Hayashi H, Nagae G, Tsuji S, Isagawa T, Midorikawa Y, Nishimura Y, Sakamoto H, Seto Y et al. 2010. Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res* **16:** 21–33.
Yamano M, Fujii H, Takagaki T, Kadowaki N, Watanabe H, Shirai T. 2000. Genetic progression and divergence in pancreatic carcinoma. *Am J Pathol* **156:** 2123–2133.
Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25:** 2865–2871.

Received March 9, 2011; accepted in revised form October 3, 2011.

nature genetics

High-resolution characterization of a hepatocellular carcinoma genome

Yasushi Totoki[1], Kenji Tatsuno[2], Shogo Yamamoto[2], Yasuhito Arai[1], Fumie Hosoda[1], Shumpei Ishikawa[3], Shuichi Tsutsumi[2], Kohtaro Sonoda[2], Hirohiko Totsuka[4], Takuya Shirakihara[1], Hiromi Sakamoto[4], Linghua Wang[2], Hidenori Ojima[5], Kazuaki Shimada[6], Tomoo Kosuge[6], Takuji Okusaka[7], Kazuto Kato[8], Jun Kusuda[9], Teruhiko Yoshida[4], Hiroyuki Aburatani[2] & Tatsuhiro Shibata[1]

Hepatocellular carcinoma, one of the most common virus-associated cancers, is the third most frequent cause of cancer-related death worldwide[1]. By massively parallel sequencing[2] of a primary hepatitis C virus–positive hepatocellular carcinoma (36× coverage) and matched lymphocytes (>28× coverage) from the same individual, we identified more than 11,000 somatic substitutions of the tumor genome that showed predominance of T>C/A>G transition and a decrease of the T>C substitution on the transcribed strand, suggesting preferential DNA repair. Gene annotation enrichment analysis[3] of 63 validated non-synonymous substitutions revealed enrichment of phosphoproteins. We further validated 22 chromosomal rearrangements, generating four fusion transcripts that had altered transcriptional regulation (*BCORL1-ELF4*) or promoter activity. Whole-exome sequencing[4,5] at a higher sequence depth (>76× coverage) revealed a *TSC1* nonsense substitution in a subpopulation of the tumor cells. This first high-resolution characterization of a virus-associated cancer genome identified previously uncharacterized mutation patterns, intra-chromosomal rearrangements and fusion genes, as well as genetic heterogeneity within the tumor.

We sequenced short-insert (250 bp, on average) genomic libraries of a primary hepatitis C virus (HCV)–positive hepatocellular carcinoma (HCC) and lymphocytes from a Japanese male (**Supplementary Fig. 1**) using the Illumina GAIIx sequencer with 50-bp paired-end reads. After alignment to the human reference genome and removal of PCR duplications, we obtained high-quality nucleotide sequences covering 102.5 Gb of the tumor genome (35.9× coverage) and 80.2 Gb (28.1× coverage) of the lymphocyte genome (**Supplementary Table 1**). The sequenced reads covered 99.69% (tumor) and 99.79% (lymphocyte) of the human reference genome. We identified 3,023,587 germline variations in the lymphocyte genome, approximately 90% of which were found in the dbSNP database, and 2,939,032 nucleotide variations in the tumor genome (a proportion of the variation was lost as a result of chromosomal alterations in the tumor genome). Comparison of the tumor and lymphocyte genomes revealed 11,731 somatically acquired nucleotide changes in the tumor genome (**Table 1**).

The prevalence of somatic substitutions was significantly less in the genic (intronic, non-coding exon and coding exon) regions relative to the intergenic regions (**Fig. 1a**, left), which could be partially explained by negative selection of lethal mutations in the gene regions or by the existence of specific molecules responsible for the repair of transcribed regions[6]. There was no significant difference in the prevalence of somatic substitutions between those of non-coding and coding exons (**Fig. 1a**, left), whereas the prevalence of germline variation was significantly decreased in the coding exons (**Fig. 1a**, right). Additionally, the ratio of non-synonymous to synonymous somatic substitutions (63/18 = 3.5) in the tumor genome was significantly higher than that of germline variations (9,573/10,552 = 0.91; $P < 0.0001$) but was not significantly different from that expected by chance (3.36; $P = 0.91$). This result suggests that an increase in negative selection of somatic substitution on the coding exons is weaker than that of germline variation. An alternative, but not mutually exclusive, explanation is that positive selection, which benefits the survival of tumor cells, partially occurs on the coding exons. The distribution of somatic substitutions revealed the dominance of T>C/A>G and C>T/G>A transitions (**Fig. 1b**). Sequence context preference was evident in some nucleotide substitutions. The C>T transition occurred significantly at CpG sites (15%; $P < 0.0001$), whereas the T>C transition occurred frequently at ApT sites (40%; $P < 0.0001$) (**Supplementary Fig. 2**). Only the T>C/A>G transition was significantly ($P = 0.01$) lower in the coding exons relative to the intergenic

[1]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [2]Genome Science Division, Research Center for Advanced Science and Technology, University of Tokyo, Meguro-ku, Tokyo, Japan. [3]Department of Pathology, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo, Japan. [4]Division of Genetics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [5]Division of Molecular Pathology, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [6]Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [7]Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [8]Institute for Research in Humanities, Graduate School of Biostudies, Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto, Japan. [9]National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan. Correspondence should be addressed to T. Shibata (tashibat@ncc.go.jp).

Received 21 July 2010; accepted 14 March 2011; published online 17 April 2011; doi:10.1038/ng.804

Table 1 Somatically acquired alterations in a liver cancer genome

Type of change	Number	Percentage
Substitutions	11,731	100.0
Coding	81	0.7
Nonsense	1	<0.1
Missense	62	0.5
Synonymous	18	0.2
Non-coding	120	1.0
UTR	83	0.7
Pseudogene	23	0.2
ncRNA	19	0.2
Intronic	4,001	34.1
Splice site	2	<0.1
Other	3,999	34.1
Intergenic	7,529	64.2
Small insertions and deletions	670	100.0
Coding	7	1.0
Non-coding	9	1.3
UTR	8	1.2
Pseudogene	0	0.0
ncRNA	2	0.3
Intronic	249	37.2
Splice site	0	0.0
Other	249	37.2
Intergenic	405	60.4
Rearrangements	22	100.0
Intrachromosomal	21	95.5
Deletions	11	50.0
Inversions	9	40.9
Tandem duplications	1	4.5
Interchromosomal	1	4.5

In 'non-coding' categories, some mutations have been classified into two subgroups. Four substitutions were classified as both UTR and non-coding RNA. One substitution was classified as both a pseudogene and non-coding RNA. One indel was classified as both UTR and non-coding RNA. UTR, untranslated region; ncRNA, non-coding RNA.

regions (**Fig. 1c**), and the C>T/G>A transition was more frequent in the coding exons relative to the intronic and non-coding exon regions, partly due to the higher GC content of coding exons and the higher frequency of CpG methylation. There were fewer T>C transitions on the transcribed strands than on the untranscribed strands ($P < 0.0001$) (**Fig. 1d**), and we observed no statistically significant differences for other substitutions.

We detected 90 somatic substitutions in protein-coding regions, 81 (including 63 non-synonymous substitutions) of which were validated as somatic alterations by Sanger sequencing of both the tumor and lymphocyte genomes (**Tables 1,2** and **Supplementary Fig. 3**). Of the remaining nine substitutions, three could not be amplified by PCR, four could not be sequenced due to the surrounding repetitive sequences and two could not be validated, likely because they were located within highly homologous segmental duplications or processed pseudogene regions. We also found evidence for 670 small somatic insertions and deletions, and all seven that are located in protein-coding regions were validated (**Tables 1** and **2, Supplementary Fig. 13**). These somatic alterations included mutations of two well-known tumor suppressor genes for HCC (*TP53* and *AXIN1*) and five genes (*ADAM22, JAK2, KHDRBS2, NEK8* and *TRRAP*) that have been found to be mutated in other cancers[7]. Gene annotation enrichment analysis[3] of the non-synonymous somatic mutations revealed significant overrepresentation of genes encoding phosphorproteins ($P = 0.0017$) and those with bipartite nuclear localization signals ($P = 0.029$) (**Supplementary Table 2**). Further re-sequencing of the exons containing potentially deleterious mutations in 96 additional pairs of primary HCC and non-cancerous liver and 21 HCC cell lines revealed two mutations (resulting in p.Phe190Leu and p.Gln212X, of which only the latter was proven to be somatic) in *LRRC30* (**Supplementary Fig. 4**). *LRRC30* contains nine repeats of a leucine-rich domain of unknown function, and all validated mutations changed the well-conserved amino acid in these repeats or produced a truncated protein.

We predicted 33 somatic rearrangements, 22 of which were validated by Sanger sequencing of the breakpoints in both the tumor and lymphocyte genomes (**Table 3**). Most of the rearrangements were intrachromosomal and occurred at the boundaries of copy number change (**Supplementary Fig. 5**). In particular, nine structural aberrations were clustered in the region of 11q12.2–11q13.4, generating a complex pattern of chromosomal amplification and loss (**Supplementary Fig. 6**). RT-PCR and sequencing analysis of the tumor and matched non-cancerous liver tissue validated four somatic fusion transcripts generated by rearrangements: the *BCORL1-ELF4* and *CTNND1-STX5* fusion genes by intra-chromosomal inversions (Xq25 and 11q12, respectively), the *VCL-ADK* fusion gene by an interstitial deletion in 10q22 (**Supplementary Fig. 7**) and the *CABP2-LOC645332* fusion gene by a tandem duplication in 11q13 (**Supplementary Fig. 8**). The *BCORL1-ELF4* chimeric transcript combining exons 1–11 of *BCORL1* and exon 8 of *ELF4* encodes an in-frame fusion protein (**Fig. 2a,b**). Quantitative RT-PCR revealed increased (>sixfold) expression of fusion transcripts in the tumor relative to wild-type *BCORL1* and *ELF4* gene expression in the non-cancerous liver (data not shown). *BCORL1* associates with CtBP and class II histone deacetylases and functions as a transcriptional repressor[8], and *ELF4* encodes a transcriptional activator[9,10] (**Fig. 2b**). We expressed *BCORL1*, *ELF4* and the chimera *BCORL1-ELF4* as Gal4-DBD fusion proteins and evaluated their transcriptional activities using a luciferase reporter assay. The chimeric protein had reduced repression activity compared to wild-type *BCORL1* (**Fig. 2c**). For the *CTNND1-STX5* fusion gene, the combination of non-coding exon 1 of *CTNND1* and exons 3–11 of *STX5* resulted in the deletion of 96 amino acids at the terminal end of *STX5* and increased (>twofold) *STX5* gene expression in the tumor,

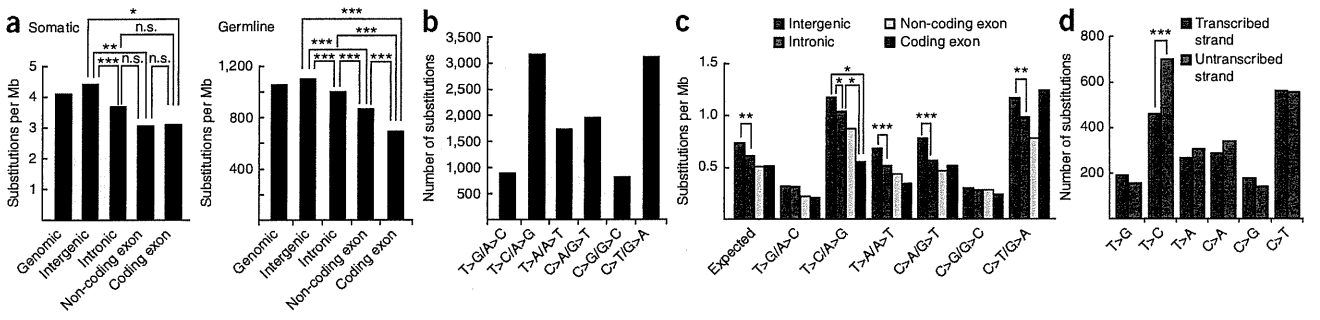

Figure 1 Somatic substitution pattern of the liver cancer genome. (**a**) Prevalence of somatic and germline substitutions in different genome regions. (**b**) Number of each type of somatic substitution in the liver cancer genome. (**c**) Prevalence of each type of somatic substitution in different genome regions. (**d**) Number of each type of somatic substitution on the transcribed and untranscribed strands. $*P < 0.05$, $**P < 0.01$, $***P < 0.0001$.

Table 2 Validated somatic non-synonymous substitutions and small indels in coding regions of a liver cancer genome

Gene	Chr.	Strand	Position	Allele change	Amino acid change	Copy number	Mutant allele (%) in whole-genome sequencing	Mutant allele (%) in whole-exome sequencing	Expression ratio (T/N)	Functional
PLEKHG5	1	–	6,452,224	G>T	Asp>Tyr	N	49.0	27.7	1.86	Deleterious
KIAA1026	1	+	15,294,007	C>A	Ala>Glu	N	45.7	nd	0.15	Tolerated
MYCL1	1	–	40,139,080	T>G	Phe>Cys	N	54.5	nd	1.93	Tolerated
PDE4B	1	+	66,231,185	C>A	Ala>Glu	N	57.1	42.9	0.83	Tolerated
CLCC1	1	–	109,284,236	A>G	Tyr>Cys	N	33.3	39.3	1.61	Deleterious
CNRIP1	2	–	68,397,833	C>T	Thr>Met	N	40.0	33.3	1.39	Deleterious
ANKRD36	2	+	97,181,397	A>G	Lys>Glu	N	17.8	nd	9.49	Tolerated
UBR3	2	+	170,511,073	A>C	Glu>Asp	N	57.1	nd	18.10	Tolerated
CUL3	2	–	225,070,790	G>A	Ser>Asn	N	42.9	52.8	12.80	Tolerated
COPS7B	2	+	232,369,129	A>G	Ile>Val	N	44.4	41.5	1.82	Tolerated
RAF1	3	–	12,625,811	A>G	Asn>Ser	N	40.0	50.0	2.31	Tolerated
ITIH3	3	+	52,813,002	A>G	Met>Val	N	43.9	nd	1.25	Deleterious
ERC2	3	–	56,148,636	G>C	Glu>Gln	N	40.0	nd	1.33	Tolerated
TBC1D23	3	+	101,496,868	del AAG	Deletion (E)	N	14.8	nd	4.90	na
ATR	3	–	143,671,657	del AT	Deletion (frame shift)	N	20.0	nd	4.49	na
SLC7A14	3	–	171,701,666	G>A	Ser>Asn	N	52.8	46.3	2.19	Deleterious
PCDH7	4	+	30,333,134	G>A	Arg>His	N	47.1	47.8	1.74	Tolerated
FAM13A	4	–	89,872,188	A>T	His>Leu	N	52.0	47.4	0.85	Tolerated
MFSD8	4	–	129,090,435	A>T	Met>Leu	Loss	62.5	74.3	1.15	Tolerated
DMGDH	5	–	78,375,996	T>A	Leu>Gln	N	50.0	37.6	3.04	Tolerated
PCDHA13	5	+	140,244,063	C>T	Pro>Ser	N	45.1	34.8	na	Deleterious
CCDC99	5	+	168,960,950	T>G	Ser>Arg	N	37.1	39.4	13.30	Deleterious
GABBR1	6	–	29,706,345	C>T	Thr>Met	N	42.0	37.8	0.59	Tolerated
CSNK2B	6	+	31,745,659	A>T	Ser>Cys	N	37.3	nd	1.41	Deleterious
MOCS1	6	–	40,003,210	G>T	Ser>Ile	N	34.4	nd	1.54	Tolerated
GTPBP2	6	–	43,699,685	A>T	Glu>Val	N	58.0	56.3	1.36	Tolerated
KHDRBS2	6	–	62,662,692	G>T	Arg>Leu	N	34.1	nd	0.88	Deleterious
SLC29A4	7	+	5,303,324	A>T	His>Leu	N	43.8	nd	7.00	Deleterious
TMEM195	7	–	15,567,887	C>G	Pro>Ala	N	41.2	38.3	1.03	Deleterious
RFC2	7	–	73,302,032	A>T	Glu>Asp	N	26.0	41.9	1.09	Tolerated
ADAM22	7	+	87,653,951	A>T	Arg>Trp	N	41.2	39.1	0.55	Deleterious
TRRAP	7	+	98,417,359	G>T	Trp>Leu	N	39.0	nd	2.07	Deleterious
XRCC2	7	–	151,977,231	G>A	Arg>Gln	N	56.2	36.5	4.18	Deleterious
MTDH	8	+	98,781,211	G>T	Val>Phe	N	33.3	46.9	14.40	Tolerated
SLA	8	–	134,141,539	C>A	Pro>Thr	N	43.6	nd	1.18	Deleterious
JAK2	9	+	5,045,703	T>G	Ile>Ser	Loss	100.0	84.2	4.84	Tolerated
NTRK2	9	+	86,532,391	G>A	Ala>Thr	Loss	90.0	85.9	0.84	Tolerated
TSC1	9	–	134,767,848	C>T	Arg>stop	Loss	13.3	13.0	1.85	Deleterious
CREM	10	+	35,496,706	A>G	Glu>Gly	N	44.8	42.3	3.28	Tolerated
C10orf95	10	–	104,200,839	T>C	Cys>Arg	N	39.7	nd	3.05	Tolerated
PSTK	10	+	124,730,061	C>T	Leu>Phe	N	53.6	nd	6.94	Deleterious
ATHL1	11	+	283,903	C>T	Ala>Val	N	40.9	26.8	1.12	Tolerated
MUC5B	11	+	1,213,214	G>T	Val>Leu	N	33.8	nd	0.83	Tolerated
DENND5A	11	–	9,181,879	C>T	Pro>Ser	N	21.4	29.9	2.43	Deleterious
GIF	11	–	59,369,438	C>T	Thr>Ile	AMP (3)	29.2	nd	0.83	Tolerated
STIP1	11	+	63,719,763	G>A	Glu>Lys	Loss	66.7	nd	1.28	Tolerated
FAT3	11	+	91,727,805	C>G	Thr>Ser	Loss	73.1	nd	na	Tolerated
PTMS	12	+	6,749,421	A>G	Glu>Gly	Loss	55.0	nd	0.56	Tolerated
ARID2	12	+	44,530,716	ins T	Insertion (frame shift)	N	31.9	nd	2.35	na
C12orf51	12	–	111,134,825	del CCTGCCACGTCA	Deletion (GDVA)	N	21.6	nd	1.44	Torelated
RBM19	12	–	112,868,641	C>T	Pro>Leu	N	49.3	42.2	1.32	Deleterious
AACS	12	+	124,142,015	G>T	Gly>Val	N	34.9	26.0	1.75	Deleterious
KHNYN	14	+	23,971,333	del CCT	Deletion (L)	N	24.1	nd	2.17	Tolerated
NOVA1	14	–	25,987,233	A>T	Leu>Phe	N	36.7	38.1	0.91	Tolerated
LTBP2	14	–	74,045,780	G>A	Gly>Glu	N	38.1	nd	3.43	Deleterious
CYFIP1	15	+	20,498,517	C>T	Ala>Val	N	55.1	41.4	1.88	Deleterious
GABRB3	15	–	24,357,328	G>T	Met>Ile	N	39.4	43.4	0.15	Tolerated
EID1	15	+	46,957,688	C>G	Ser>Cys	N	40.4	nd	8.60	Deleterious
HCN4	15	–	71,402,254	G>A	Arg>His	N	43.6	nd	0.61	Tolerated
AKAP13	15	+	84,060,152	del T	Deletion (frame shift)	N	34.5	nd	0.88	na
AXIN1	16	–	287,910	C>T	Arg>stop	Loss	78.7	nd	0.94	Deleterious
LITAF	16	–	11,554,943	del G	Deletion (frame shift)	Loss	61.3	nd	0.97	na
TP53	17	–	7,518,985	G>T	Val>Leu	Loss	78.0	73.1	0.06	Deleterious
NEK8	17	+	24,092,271	G>A	Gly>Asp	N	36.7	39.1	1.44	Deleterious
CPD	17	+	25,773,820	A>G	Tyr>Cys	N	47.1	52.3	2.28	Deleterious
LRRC30	18	+	7,221,594	C>G	Ser>Cys	N	52.0	45.6	na	Deleterious
ZNF560	19	–	9,439,794	A>C	Ile>Leu	N	58.8	48.3	0.86	Tolerated
SCRT2	20	–	593,073	T>A	Tyr>Asn	N	53.7	nd	0.51	Deleterious
USP25	21	+	16,119,227	C>T	Thr>Met	N	44.4	nd	13.00	Deleterious
USP25	21	+	16,125,626	A>C	Glu>Asp	N	35.3	38.1	na	Tolerated
ARVCF	22	–	18,341,717	C>G	Ser>Cys	N	53.0	50.0	1.30	Deleterious
USP26	X	–	131,988,824	T>C	Leu>Pro	AMP (4)	93.8	94.4	0.85	Tolerated

Except for *ANKRD36* and *TSC1*, all 63 somatic non-synonymous substitutions were predicted by whole-genome sequencing and in-house informatics method using stringent analysis criteria (Online Methods). One somatic missense substitution in *ANKRD36* was predicted under less stringent criteria. One somatic nonsense substitution in *TSC1* was predicted only by whole-exome sequencing. Chr., chromosome; N, copy neutral; AMP, amplicon; nd, not detected; na, not applicable.

Table 3 Validated somatic structural alterations in a liver cancer genome

Type	Chr. A	Break point A	CNV (Chr. A)	Chr. B	Break point B	CNV (Chr. B)	Intervening sequence	Associated genes	Fusion genes
Deletion	3	111,866,468	BCNC	3	111,868,894	BCNC	0		
Deletion	4	57,529,004	BCNC	4	57,530,452	BCNC	0	*C4orf14* (exon 4 is deleted)	
Deletion	4	92,895,135	BCNC	4	93,151,201	BCNC	0		
Deletion	5	18,130,563	BCNC	5	18,133,946	BCNC	(+) 29bp		
Deletion	6	90,130,109	BCNC	6	90,819,100	BCNC	0	*LYRM2, ANKRD6, BACH2, MDN1, CASP8AP2, RRAGD, GJA10*	
Deletion	7	69,321,043	N	7	69,404,639	N	0	*AUTS2*	
Deletion	9	132,763,157	BCNC	9	132,764,920	BCNC	0		
Deletion	10	75,477,784	BCNC	10	75,956,310	BCNC	(+) 1 bp	*AP3M1, VCL, ADK*	*VCL, ADK*
Deletion	11	67,126,436	BCNC	11	68,254,241	BCNC	0	*SUV420H1, SAPS3, ACY3, ALDH3B2, CHKA, TCIRG1, LRP5, GAL, ALDH3B1, TBX10, NDUFV1, UNC93B1, NUDT8, C11orf24*	
Deletion	15	47,394,203	BCNC	15	47,467,920	BCNC	0	*GALK2, C15orf33*	
Deletion	17	15,902,440	BCNC	17	16,056,159	BCNC	0	*NCOR1* (homozygous deletion)	
Inversion	4	60,946,299	N	4	60,947,151	N	0		
Inversion	4	172,703,199	Loss	4	172,706,239	Loss	(+) 4bp		
Inversion	11	57,305,269	BCNC	11	62,352,275	BCNC	0	*CTNND1* (UTR), *STX5*	*CTNND1, STX5*
Inversion	11	57,770,822	BCNC	11	67,133,985	BCNC	0	*NDUFV1*	
Inversion	11	62,309,952	BCNC	11	70,746,006	BCNC	0	*TAF6L*	
Inversion	11	69,067,231	AMP	11	69,317,424	AMP	0		
Inversion	11	69,093,978	AMP	11	69,098,117	AMP	0		
Inversion	11	69,871,206	AMP	11	69,877,391	AMP	(+) 6bp	*PPFIA1*	
Inversion	X	129,015,072	N	X	129,029,501	BCNC	(+) 23bp	*BCORL1, ELF4*	*BCORL1, ELF4*
Inversion	X	129,016,981	N	X	129,031,425	BCNC	0	*BCORL1, ELF4*	*BCORL1, ELF4*
Tandem duplication	11	67,043,308	BCNC	11	67,318,685	BCNC	0	*ACY3, ALDH3B2, GSTP1, TBX10, NDUFV1, NUDT8, CABP2, LOC645332*	*CABP2, LOC645332*
Translocation	11	69,316,960	AMP	X	129,030,346	BCNC	0	*ELF4*	

The inversions at Xq25 occurred from one rearrangement event and the total number of inversion is counted as nine. Chr., chromosome; BCNC, boundary of copy number change; N, copy neutral; AMP, amplicon.

which harbors only the rearranged allele (**Fig. 2d** and **Supplementary Fig. 9**). We screened for the presence of these four chimera transcripts by RT-PCR, but we detected no recurrent fusion event in 47 cases of primary HCC, possibly due to the low frequency of these rearrangements in HCC or because of the technical difficulty in detecting all variant fusion transcripts.

We also sequenced the whole exomes of the same samples using an in-solution gene enrichment system[5] (**Fig. 3a**). Capture probes for whole-exome sequencing were designed to cover the protein coding exons using the consensus coding sequences, excluding highly homologous regions. The average coverage of the whole exome sequences (41.3 Mb in total) was about twice (76.8× for HCC and 74.3× for lymphocytes) that of the whole genome sequences and had one twelfth of the total sequence amount (8.9 Gb for HCC and 8.6 Gb for lymphocyte) (**Supplementary Table 3**). Whole-exome sequencing detected 47 non-synonymous somatic substitutions, 40 of which were validated by Sanger sequencing. Among the validated substitutions, a nonsense substitution (p.Arg785X) in *TSC1*, located in the hemizygous region (9q34), was not detected by whole-genome sequencing (**Fig. 3b**). Capillary sequencing validated the same substitution with a very low

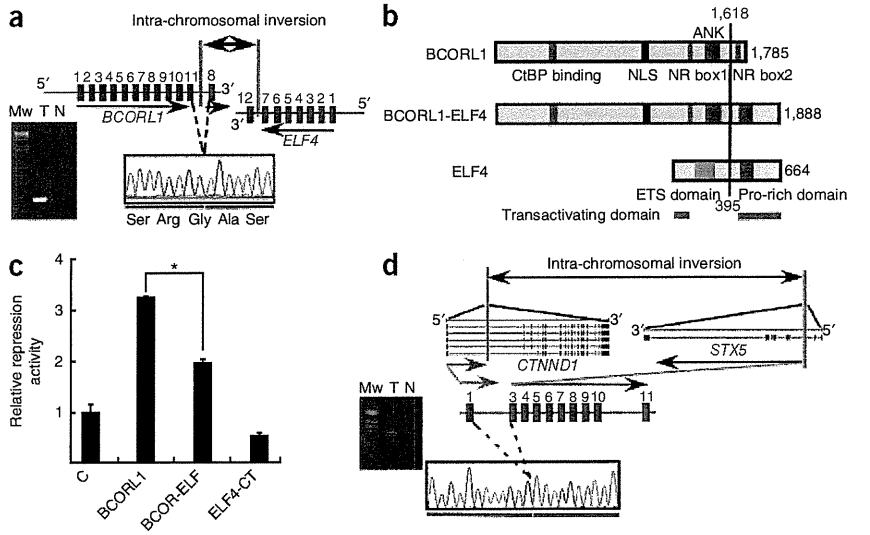

Figure 2 Characterization of rearrangements in liver cancer. (**a**) Top, schematic representation of the intra-chromosomal inversion at Xq25. Bottom left, RT-PCR analysis of the fused *BCORL1-ELF4* transcript in tumor (T) and non-cancerous liver (N) tissues. We detected no *ELF4-BCORL1* transcript (data not shown). Bottom right, sequence chromatography of the fusion transcript revealed an in-frame protein. Mw, molecular marker. (**b**) Schematic representation of the BCORL1-ELF4 fusion protein. BCORL1 (top) contains a CtBP1 binding domain (PXDLS sequence), a binuclear localization signal (NLS), two LXXLL nuclear receptor recruitment motifs (NR box) and tandem ankyrin repeats (ANK). ELF4 (bottom) contains an ETS (E Twenty Six) DNA binding domain and a proline-rich domain. Transactivating domains are indicated by the red bars[16]. The BCORL1-ELF4 chimeric protein includes most of BCORL1 (1–1,618 amino acids) lacking the NR box2 and the carboxyl-terminal portion of ELF4 containing the proline-rich domain. The number of amino acids is indicated on the right. (**c**) Wild-type BCORL1, ELF4-CT (395–664 amino acids) and the BCORL1-ELF4 chimera were expressed as Gal4-DBD fusion proteins, and their relative transcriptional activities were compared to the Gal4-DBD protein (C) as shown. (**d**) Characterization of the *CTNND1-STX5* fusion gene. Bottom left, RTPCR analysis of the fused *CTNND1-STX5* transcript in tumor (T) and non-cancerous liver tissue (N). Bottom right, sequence chromatography of the fusion transcript. Data is the mean ± s.d. ($n = 3$). *$P < 0.001$.

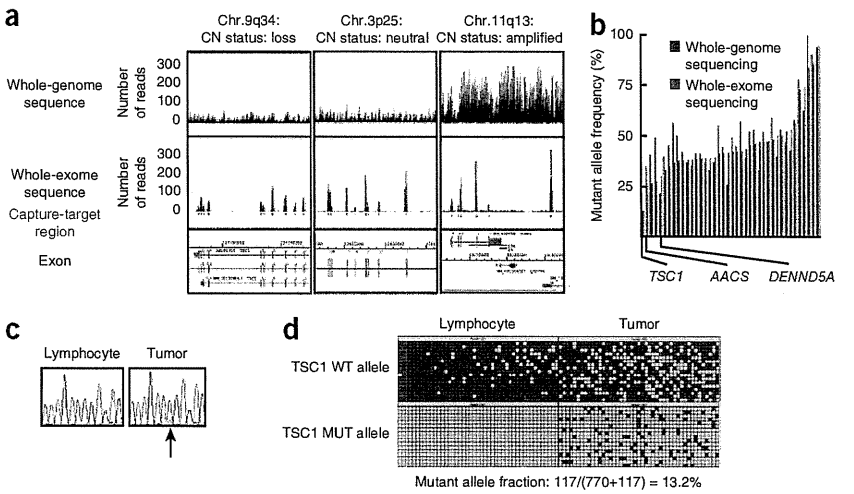

Figure 3 Intra-tumoral genetic heterogeneity detected by exon-capture sequencing. (**a**) Specific enrichment and high sequence coverage of the target genome regions indicated by the sequence viewer (copy number (CN) status is shown above). The distribution and number of reads (black, forward read; gray, reverse read) from whole-genome sequencing (top) and whole-exome sequencing (middle) are shown. The location of the capture target regions (red box) and the exons (green box) along the genome are shown at the bottom. Note that the number of reads is dependent on copy number status. (**b**) Mutant allele frequency detected by whole-genome sequencing and whole-exome sequencing. *TSC1*, *AACS* (whose heterogeneity is shown in **Supplementary Fig. 10**) and *DENND5A* are indicated. (**c**) *TSC1* mutation in the liver cancer subpopulation. Sequence chromatograms of *TSC1* in lymphocytes and whole-tumor tissue are shown. Note the small peak for the mutant T allele (indicated by the arrow) in the tumor DNA. (**d**) Determination of mutant *TSC1* allele frequency by digital PCR genotyping. WT, wild type; MUT, mutant.

signal peak (**Fig. 3c**), and digital genotyping showed that 13.2% of the tumor alleles harbored this substitution (**Fig. 3d**), suggesting that this substitution occurred in a minor population of cancer cells. Whole-exome sequencing missed 25 non-synonymous somatic substitutions that were detected by whole-genome sequencing. These missed substitutions were located in regions where sequence coverage was low or where further optimization of the probe design was required.

The number of non-synonymous somatic substitutions validated in this HCC (63) was greater than those for acute myeloid leukemia[11] (10), basal-like breast cancer[12] (22), lobular carcinoma[13] (32), glioblastoma multiforme[14] (32) and pancreatic cancer[15] (43) but is in the range of those previously reported for colorectal[16] (70) and breast[16] (88) cancer. We have shown that the pattern of somatic substitutions in a HCV-associated HCC genome is different (predominance of T>C, especially at ApT sites, and C>T, especially at CpG sites) compared to smoking-related[17,18] and ultraviolet light–related[6] cancers. Preferential C>T/G>A transition may partly be due to the higher frequency of CpG methylation in the genome sequence and is a common form of mutation in cancers[19]. Therefore, the T>C/A>G transition could be a characteristic mutational signature of HCV-associated cancer, which would be consistent with a previous observation that HCV induces error-prone DNA polymerases that preferentially cause the T>C/A>G mutation[20]. It is also possible that this mutation pattern is independent of viral infection and is organ specific, as a comparable substitution spectrum has been reported in renal cancer[19]. Additionally, only T>C changes, but not C>T changes, were effectively repaired on the transcribed strand. Similar enhanced transcription-coupled repair on preferentially acquired substitutions has been reported in other cancers[6,17,18] and could be a common phenomenon in cancer mutation.

Because single-molecule sequencing has the capability to detect every individual somatic event in parallel, higher sequence coverage will enable us to clarify the intra-tumoral heterogeneity that is associated with diverse aspects of clinical behavior such as metastasis[21]. The *TSC1* complex, which is inactivated in a subpopulation of tumors, negatively regulates the mammalian target of rapamycin signaling, which is an important oncogenic pathway related to the growth, metabolism and stemness of cancer cells[22,23], and could be a promising molecular therapeutic target in HCC progression[24].

URLs. International Cancer Genome Consortium, http://www.icgc.org/; Catalogue of Somatic Mutations in Cancer, http://www.sanger.ac.uk/genetics/CGP/cosmic/; BLASTN, ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/LATEST.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank K.K. Khanna (The Queensland Institute of Medical Research) for providing a human *BCORL1* cDNA clone; T.D. Taylor (RIKEN) for comments on the manuscript; T. Urushidate, S. Ohashi, S. Ohnami, A. Kokubu, N. Okada, K. Shiina, H. Meguro and K. Nakano for their excellent technical assistance. This work was supported by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO), Japan, and the Industrial Technology Research Grant Program from the New Energy and Industrial Technology Development Organization (NEDO), Japan. This study is associated with the International Cancer Genome Consortium (ICGC), and the mutation data were deposited at and released from the ICGC web site.

AUTHOR CONTRIBUTIONS

The study was designed by T. Shibata, H.A., T.Y. and J.K. Sequencing and data analyses were conducted by Y.T., K.T., S.Y., S.T., K. Sonoda and H.T. Allele typing and copy number analyses were performed by H.S. and S.I. Other molecular studies were done by Y.A., F.H., T. Shirakihara, and L.W.; H.O., K. Shimada, T.K., T.O. and K.K. coordinated collection of clinical sample and information. The manuscript was written by Y.T., T. Shibata, K.T., S.Y., H.A. and T.Y.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics /.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/ .

1. El-Serag, H.B. & Rudolph, K.L. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* **132**, 2557–2576 (2007).
2. Bentley, D.R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
3. Huang, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nat. Protoc.* **4**, 44–57 (2009).
4. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).

5. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
6. Pleasance, E.D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
7. Xiang, Z. *et al.* Identification of somatic *JAK1* mutations in patients with acute myeloid leukemia. *Blood* **111**, 4809–4812 (2008).
8. Pagan, J.K. *et al.* A novel corepressor, BCoR-L1, represses transcription through an interaction with CtBP. *J. Biol. Chem.* **282**, 15248–15257 (2002).
9. Miyazaki, Y., Sun, X., Uchida, H., Zhang, J. & Nimer, S. MEF, a novel transcription factor with an Elf-1 like DNA binding domain but distinct transcriptional activating properties. *Oncogene* **13**, 1721–1729 (1996).
10. Suico, M.A. *et al.* Functional dissection of the ETS transcription factor MEF. *Biochim. Biophys. Acta* **1577**, 113–120 (2002).
11. Mardis, E.R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
12. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
13. Shah, S.P. *et al.* Mutation evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
14. Parsons, D.W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
15. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
16. Wood, L.D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
17. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
18. Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
19. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
20. Machida, K. *et al.* Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. *Proc. Natl. Acad. Sci. USA* **101**, 4262–4267 (2004).
21. Kim, M.Y. *et al.* Tumor self-seeding by circulating cancer cells. *Cell* **139**, 1315–1326 (2009).
22. Guertin, D.A. & Sabatini, D.M. Defining the role of mTOR in cancer. *Cancer Cell* **12**, 9–22 (2007).
23. Yilmaz, O.H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
24. Meric-Bernstam, F. & Gonzalez-Angulo, A.M. Targeting the mTOR signaling network for cancer therapy. *J. Clin. Oncol.* **27**, 2278–2287 (2009).

ONLINE METHODS

Whole-genome sequencing. High molecular weight DNA was extracted from freshly frozen tumor tissue and lymphocytes. DNA was fragmented using an ultrasonic solubilizer (Covaris) using a combination of quick bursts (20% duty, 5 intensity with 200 cycles per burst for 5 s) and sonication (10% duty, 5 intensity with 200 cycles per burst for 120 s) for the short fragment DNA library. DNA of the appropriate size was gel purified to exclude any inappropriate DNA fusions during library construction. The short fragment DNA libraries were generated using a paired-end DNA sample prep kit (Illumina) following the manufacturer's protocols. The concentration of the libraries was quantified using a Bioanalyzer (Agilent Technologies); 4~8 pM/lane of DNA was applied to the flow cell, and paired-end sequencing was performed using the GAIIx sequencer (Illumina).

Whole-exome capture sequencing. Whole-exome capture sequencing was performed using the SureSelect Target Enrichment System (Agilent Technologies) in accordance with the manufacturer's protocol with slight modifications. Briefly, the same Illumina sequence libraries as those prepared for the whole-genome sequence were amplified with six cycles of PCR, and then 500 ng of the amplified libraries was hybridized with the capture probes for 24 h. The hybridized sequence libraries were collected and further amplified with 14 cycles of PCR. We generated 51-nucleotide–long paired-end reads using the GAIIx sequencer (Illumina). We used five lanes of a paired-end flow cell for each sample.

Bioinformatics (Supplementary Fig. 11). *Sequence alignment to the human genome and removal of PCR duplications.* Paired-end reads were aligned to the human reference genome (hg18, NCBI Build 36.1) using Burrows-Wheeler Aligner (BWA) (version 0.4.9)[25]. Because there were duplicated reads which were generated during the PCR amplification process, paired-end reads that aligned to the same genomic positions were removed using SAMtools (version 0.1.5c)[26] and a program developed in house. We removed 12.5% (14.6/117.1 Gbp) of the aligned reads for tumor and 7.1% (6.1/86.3 Gbp) for lymphocytes.

Detection of somatic single nucleotide variations (SNVs) **(Supplementary Fig. 12).** Based on the genotyping data from two SNP arrays, appropriate thresholds for base quality, mapping quality and frequency of non-reference alleles were determined to obtain the highest confidence calls for SNV detection **(Supplementary Table 4).** To predict somatic SNVs, the alignment results were classified, and three datasets were constructed. Dataset 1 included paired-end reads with both ends aligned uniquely and with proper spacing and orientation. Dataset 2 included paired-end reads that aligned uniquely for at least one read and with proper spacing and orientation of the reads. Dataset 3 included dataset 2 and paired-end reads for which both ends aligned uniquely but with improper spacing or orientation or both. Dataset 1 likely contains false positive somatic SNVs because of the low sequence depth of the lymphocyte genome, and dataset 3 likely contains false positives due to misalignments of the sequence reads. To reduce the number of false positives, the following filters were applied to these three datasets, and concordant somatic SNVs among the three datasets were selected: (i) a mapping quality score of 20 was used as a cutoff value for read selection; (ii) base quality scores of 10 and 15 were used as cutoff values for base selection for the tumor and lymphocyte genomes, respectively; (iii) SNVs were selected when the frequency of the non-reference allele was at least 15% in the tumor genome and 5% in the lymphocyte genome; (iv) SNVs located within 5 bp from a potential insertion or deletion were discarded; (v) SNVs with a root mean square mapping quality score of the reads covering the SNV less than 40 were discarded; (vi) when there were three or more SNVs within any 10-bp window, all of them were discarded; (vii) SNVs with a consensus quality score less than 20 as calculated by SAMtools (version 0.1.5c) were discarded; (viii) when a base with a consensus quality score less than 20 was located within 3-bp on either side of a SNV, the SNV was discarded; (ix) for the tumor genome, SNVs found in at least two sequence reads with the same SNV were selected; (x) for the lymphocyte genome, SNVs covered by at least six sequence reads were selected; and (xi) the repetitive regions within 1 Mb of a centromeric or telomeric sequence gap were excluded. By comparing the predicted nucleotide variations in the tumor and lymphocyte genomes, somatic SNVs which occurred only in the tumor genome were identified. If somatic SNVs were not covered in the lymphocyte genome by at least six sequence reads, they were discarded.

Using this approach, 66 non-synonymous and 24 synonymous somatic SNVs in protein-coding regions were predicted. These 90 substitutions were examined by Sanger sequencing of both the tumor and lymphocyte genomes, and 81 of them were validated as somatic mutations. Of the remaining nine substitutions, three could not be amplified by PCR, four could not be sequenced because of the surrounding repetitive sequences, and two could not be validated likely because they were located in highly homologous segmentally duplicated or processed pseudogene regions, suggesting a high prediction accuracy (specificity, 81/83 = 97.6%) for our approach for detecting somatic SNVs in protein-coding regions. An additional 36 non-synonymous somatic SNVs were also predicted using only dataset 3 and filtering methods (i–iv) (less stringent filtering condition). Five of these SNVs were not validated and 30 of them were found to be germline variations by Sanger sequencing, and only the one remaining was validated as a somatic mutation. These findings suggest that our filtering method (stringent condition) effectively removed false-positive somatic SNVs.

Detection of somatic structural alterations. To detect structural alterations, paired-end reads for which both ends aligned uniquely to the human reference genome, but with improper spacing or orientation or both, were used. First, paired-end reads were selected based on the following filtering conditions: (i) sequence reads with mapping quality scores greater than 37; and (ii) sequence reads aligned with two mismatches or less.

Rearrangements were then identified using the following analytical conditions: (i) 'clusters' which included reads aligned within the maximum insert distance were constructed from the forward and reverse alignments, respectively (two reads were allocated to the same cluster if their end positions were not further apart than the maximum insert distance); (ii) clusters whose distance between the leftmost and rightmost reads were greater than the maximum insert distance were discarded; (iii) paired-end reads were selected if one end sequence was allocated in the 'forward cluster' and the other end was allocated in the 'reverse cluster' (we called these 'forward cluster and reverse cluster' paired clusters); (iv) if a cluster overlapped another cluster, all of the overlapping paired-clusters were discarded; (v) for the tumor genome, rearrangements (paired-clusters) predicted by at least four paired-end reads which included at least one paired-end read perfectly matched to the human reference genome were selected; and (vi) for the lymphocyte genome, rearrangements (paired clusters) predicted by at least one paired-end read were selected. By comparing the predicted rearrangements in the tumor and lymphocyte genomes, somatic rearrangements that were only detected in the tumor genome were identified.

Lastly, rearrangements predicted due to variations in the analyzed genomes were removed. For this analysis, paired-end reads contained in paired clusters were aligned to the human reference genome using the BLASTN program (see URLs). If one end sequence was aligned to the region of paired clusters (the flanking region of the rearrangement breakpoint) and the other end was aligned with proper spacing and orientation, the rearrangement was removed. An expectation value of 1,000 was used as a cutoff value for BLASTN so that paired-end reads with low similarity to the human reference genome could also be aligned.

Using this method, 33 somatic rearrangements were predicted and 22 of these were validated by Sanger sequencing of the rearrangement breakpoints in both the tumor and lymphocyte genomes.

Exome capture sequence analysis. To analyze the capture sequencing data, the Illumina sequencing pipeline version 1.4 and in-house programs were used. The sequence reads were mapped to the human reference sequence (NCBI Build 36.3) using GERALD (Illumina), and only high-quality ('pass filter') reads with base-call quality scores more than ten were used for SNV detection.

SNVs were determined using the frequency (>20%) of the highest non-reference base call with a read depth greater than 20×.

Other molecular analyses. SNP genotyping and copy number detection were determined using the Affymetrix Mapping 500K Array, the Agilent Human Genome CGH microarray and the Illumina Human 610-Quad BeadChip system. Gene expression levels of the tumor were measured using the Agilent Whole Human Genome Oligo Microarray. Wild-type and mutant allele frequencies were determined using the Digital PCR system.

Detailed experimental methods and additional bioinformatics procedures are described in **Supplementary Note**. The somatic substitutions and insertions/deletions found are listed in **Supplementary Tables 5–9**.

25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754–1760 (2009).
26. Li, H. *et al.* The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

OPEN ACCESS Freely available online

PLoS one

Increased Systemic Glucose Tolerance with Increased Muscle Glucose Uptake in Transgenic Mice Overexpressing RXRγ in Skeletal Muscle

Satoshi Sugita[1], Yasutomi Kamei[1]*, Fumiko Akaike[1,2], Takayoshi Suganami[1], Sayaka Kanai[1], Maki Hattori[1,2], Yasuko Manabe[3], Nobuharu Fujii[3], Takako Takai-Igarashi[4], Miki Tadaishi[5,6], Jun-Ichiro Oka[2], Hiroyuki Aburatani[7], Tetsuya Yamada[8], Hideki Katagiri[8], Saori Kakehi[9], Yoshifumi Tamura[9,10], Hideo Kubo[11], Kenichi Nishida[11], Shinji Miura[5], Osamu Ezaki[5], Yoshihiro Ogawa[1,12]

1 Department of Molecular Medicine and Metabolism, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan, **2** Laboratory of Pharmacology, Faculty of Pharmaceutical Sciences, Tokyo University of Science, Chiba, Japan, **3** Graduate School of Human Health Sciences, Tokyo Metropolitan University, Tokyo, Japan, **4** Department of Bioinformatics, Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan, **5** Nutritional Science Program, National Institute of Health and Nutrition, Tokyo, Japan, **6** Department of Nutritional Science, Faculty of Applied Bioscience, Tokyo University of Agriculture, Tokyo, Japan, **7** Research Center for Advanced Science and Technology, University of Tokyo, Tokyo, Japan, **8** Department of Metabolic Diseases, Center for Metabolic Diseases, Tohoku University Graduate School of Medicine, Miyagi, Japan, **9** Department of Medicine, Metabolism and Endocrinology, School of Medicine, Juntendo University, Tokyo, Japan, **10** Sportology Center, Juntendo University, Tokyo, Japan, **11** Daiichi-Sankyo Co., Ltd., Tokyo, Japan, **12** Global Center of Excellence Program, International Research Center for Molecular Science in Tooth and Bone Diseases, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

Abstract

Background: Retinoid X receptor (RXR) γ is a nuclear receptor-type transcription factor expressed mostly in skeletal muscle, and regulated by nutritional conditions. Previously, we established transgenic mice overexpressing RXRγ in skeletal muscle (RXRγ mice), which showed lower blood glucose than the control mice. Here we investigated their glucose metabolism.

Methodology/Principal Findings: RXRγ mice were subjected to glucose and insulin tolerance tests, and glucose transporter expression levels, hyperinsulinemic-euglycemic clamp and glucose uptake were analyzed. Microarray and bioinformatics analyses were done. The glucose tolerance test revealed higher glucose disposal in RXRγ mice than in control mice, but insulin tolerance test revealed no difference in the insulin-induced hypoglycemic response. In the hyperinsulinemic-euglycemic clamp study, the basal glucose disposal rate was higher in RXRγ mice than in control mice, indicating an insulin-independent increase in glucose uptake. There was no difference in the rate of glucose infusion needed to maintain euglycemia (glucose infusion rate) between the RXRγ and control mice, which is consistent with the result of the insulin tolerance test. Skeletal muscle from RXRγ mice showed increased Glut1 expression, with increased glucose uptake, in an insulin-independent manner. Moreover, we performed *in vivo* luciferase reporter analysis using *Glut1* promoter (*Glut1*-Luc). Combination of RXRγ and PPARδ resulted in an increase in *Glut1*-Luc activity in skeletal muscle *in vivo*. Microarray data showed that RXRγ overexpression increased a diverse set of genes, including glucose metabolism genes, whose promoter contained putative PPAR-binding motifs.

Conclusions/Significance: Systemic glucose metabolism was increased in transgenic mice overexpressing RXRγ. The enhanced glucose tolerance in RXRγ mice may be mediated at least in part by increased Glut1 in skeletal muscle. These results show the importance of skeletal muscle gene regulation in systemic glucose metabolism. Increasing RXRγ expression may be a novel therapeutic strategy against type 2 diabetes.

Citation: Sugita S, Kamei Y, Akaike F, Suganami T, Kanai S, et al. (2011) Increased Systemic Glucose Tolerance with Increased Muscle Glucose Uptake in Transgenic Mice Overexpressing RXRγ in Skeletal Muscle. PLoS ONE 6(5): e20467. doi:10.1371/journal.pone.0020467

Editor: Massimo Federici, University of Tor Vergata, Italy

Received March 15, 2011; **Accepted** April 26, 2011; **Published** May 31, 2011

Funding: This work was supported in part by a Grant-in-Aid for scientific research KAKENHI from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT, Tokyo, Japan), research grants from the Japanese Ministry of Health, Labor and Welfare, and by the joint research program of the Institute for Molecular and Cellular Regulation, Gunma University. S. Sugita is a Research Fellow of the Japan Society for the Promotion of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: HK and KN are currently employed by the commercial company Daiichi-Sankyo Co., Ltd. HK and KN were non-commercially involved in this study. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials. Therefore, HK and KN do not have competing interests to disclose.

* E-mail: kamei.mmm@mri.tmd.ac.jp

Introduction

The skeletal muscle, known as the largest organ in the human body, plays an important role in exercise, energy expenditure, and glucose metabolism. It is a major site of glucose disposal [1,2]. In type 2 diabetic subjects, glucose uptake in the skeletal muscle is impaired [2]. Blood glucose is taken up by the skeletal muscle via insulin-dependent and independent glucose transporters (Glut4

and Glut1, respectively) [3], where it is converted into glycogen. Increasing capacity of glucose uptake in the skeletal muscle is considered beneficial for the prevention and treatment of type 2 diabetes [3,4]. On the other hand, glucose metabolism in the skeletal muscle may affect the whole body metabolism; the skeletal muscle-specific inactivation of *Glut4* has resulted in defect in insulin action in the adipose tissue (glucose uptake) and liver (suppression of gluconeogenesis) [5]. Thus, identification of the molecular mechanisms involved in skeletal muscle glucose metabolism should help clarify the pathophysiology of diabetes.

Nuclear receptors are part of a large superfamily of transcription factors that includes receptors for steroids, retinoic acid, and thyroid hormones [6]. RXRs are heterodimeric partners of many nuclear receptors, such as retinoic acid receptors (RARs), thyroid hormone receptors (T3Rs), liver X receptors (LXRs), peroxisome proliferator activated receptors (PPARs), and RXRs themselves [6]. Among RXR heterodimer partners, activation of PPARδ in the skeletal muscle increases insulin sensitivity [7]. The RXR subfamily consists of RXRα, RXRβ, and RXRγ [6,8]. Although RXRγ is preferentially expressed in the skeletal muscle, its functional role is poorly understood. We have recently found that expression of retinoid X receptor γ (RXRγ) is changed in the skeletal muscle under nutritional conditions; RXRγ mRNA expression is down-regulated by fasting and recovered by refeeding [9]. In an attempt to explore the role of RXRγ in the skeletal muscle, we established transgenic mice overexpressing RXRγ in the skeletal muscle (RXRγ mice) and found that they exhibit increased triglyceride contents in the skeletal muscle as a result of increased expression of sterol regulatory element binding protein 1c (SREBP1c), a transcriptional master regulator of lipogenesis [9]. Indeed, RXRγ has been shown to enhance SREBP1c gene expression in C2C12 myocytes *in vitro* at least in part by heterodimerization with LXR [9]. On the other hand, we also found that blood glucose levels are lower in RXRγ mice than in control mice [9]. These observations, taken together, suggest that RXRγ plays a critical role in glucose and lipid metabolism in the skeletal muscle. However, the molecular mechanism involved in RXRγ regulation of glucose metabolism in the skeletal muscle and how it affects systemic glucose metabolism are poorly understood.

Here, we demonstrate enhanced glucose metabolism with increased Glut1 expression and glucose uptake in RXRγ mice. This study suggests that activation of the skeletal muscle RXRγ is a novel therapeutic strategy to treat or prevent type 2 diabetes.

Methods

Animals

C57BL6 mice were purchased from Charles River Japan (Yokohama, Japan). Generation of RXRγ mice under the control of the human α-actin promoter was described previously [9]. They were allowed free access to food (CRF-1; Charles River) and water, unless otherwise stated. All animal experiments were approved by Institutional Animal Care and Use Committee of Tokyo Medical and Dental University (approval ID: No. 0090041).

Blood analysis

Serum samples were obtained from mice, when fed *ad libitum*. Serum glucose levels were measured by the blood glucose test meter (Glutest PRO R; Sanwa-Kagaku, Nagoya, Japan). Serum concentrations of insulin were determined by the enzyme-linked immunosorbent assay (ELISA) kits (Morinaga Institute of Biological Science, Inc., Yokohama, Japan).

Glucose and insulin tolerance tests

For glucose tolerance test, D-glucose (1 mg/g of body weight, 10% (w/v) glucose solution) was administered by intraperitoneal injection after an overnight fast. For insulin tolerance test, human insulin (Humulin R; Eli Lilly Japan K.K., Kobe, Japan) was injected intraperitoneally (0.75 mU/g of body weight), when fed *ad libitum*.

Hyperinsulinemic-euglycemic clamp studies

We analyzed as described previously [10–12] with slight modification. Two days before the clamp studies, a catheter was inserted into the right jugular vein for infusion under general anesthesia with sodium pentobarbital. Studies were performed on mice under conscious and unstressed conditions after a 4-h fast. The clamp study began with a prime (4 mg/kg for 5 min) of [6,6-^{2}H] glucose (Cambridge Isotope Laboratories, Inc, Andover, MA) followed by continuous infusion at a rate of 0.5 mg/kg per minute for 2 hr to assess the basal glucose turnover. After the basal period, hyperinsulinemic-euglycemic clamp was conducted for 120 min with a primed/continuous infusion of human insulin (25 mU/kg prime for 5 min, 5 mU/kg/min infusion) and variable infusion of 20% glucose to maintain euglycemia (approximately 100 mg/dl). The 20% glucose was enriched with [6,6-^{2}H] glucose to approximately 2.5% as previously described [12]. To determine the enrichment of [6,6-^{2}H]glucose in plasma at basal and insulin stimulated state, samples were deproteinized with trichloroacetic acid and derivatized with *p*-aminobenzoic acid ethyl ester. The atom percentage enrichment of glucose$_{m+2}$ was then measured by high-performance liquid chromatography with LTQ-XL-Orbitrap mass spectrometer (Thermo Scientific, CA). The glucose$_{m+2}$ enrichment was determined from the *m/z* ratio 332.2: 330.2. The hepatic glucose production was calculated by using the rate of infusion of [6,6-^{2}H]glucose over the atom percent excess in the plasma minus the rate of glucose being infused. The insulin-stimulated whole-body glucose uptake was calculated by adding the total glucose infusion rate plus the hepatic glucose production [12].

Quantitative real-time PCR. Quantitative real-time PCR was performed as described [9]. Total RNA was prepared using Sepazol (Nacalai Tesque, Kyoto, Japan). cDNA was synthesized from 5 μg of total RNA using Superscript II reverse transcriptase (Invitrogen Inc., Carlsbad, CA) with random primers. Gene expression levels were measured with an ABI PRISM 7700 using SYBR Green PCR Core Reagents (Applied Biosystems, Tokyo, Japan). The primers used were as follows, RXRγ: Fw: 5′-CAC-

Table 1. Body and dissected tissue weight and blood glucose level in RXRγ mice (4-3 line).

	Wild-type	RXRγ
Body weight (g)	26.4±0.3	26.5±0.4
Epididymal fat mass (g)	0.138±0.004	0.156±0.090
Gastrocnemius muscle weight (g)	0.282±0.004	0.238±0.008**
Liver weight (g)	1.290±0.029	1.321±0.033
Glucose (mg/dL) basal	170.0±9.4	145.6±4.9*
Glucose (mg/dL) fasting	74.4±1.7	67.0±0.6**

Mice were males 12 weeks of age. The number of animals used was 6 for both wild-type control and RXRγ mice.
* $P<0.05$;
** $P<0.01$, compared with wild-type control.
Values are the means ± SE. These samples were also used in Fig 2*A*, *B*, and *D*.
doi:10.1371/journal.pone.0020467.t001

Table 2. Body and dissected tissue weight and blood glucose level in RXRγ mice (5-3 line).

	Wild-type	RXRγ
Body weight (g)	36.0±1.8	31.9±1.2
Epididymal fat mass (g)	1.030±0.191	0.861±0.134
Gastrocnemius muscle weight (g)	0.305±0.016	0.214±0.020**
Liver weight (g)	1.375±0.119	1.320±0.066
Glucose (mg/dL) basal	142.5±8.6	117.3±8.2
Glucose (mg/dL) fasting	67.2±1.6	53.3±2.5**

Mice were males 38 weeks of age. The number of animals used was 6 for both wild-type control and RXRγ mice.
** $P<0.01$, compared with wild-type control.
Values are the means ± SE.
doi:10.1371/journal.pone.0020467.t002

CCTGGAGGCCTATACCA-3′, Rv: 5′-AAACCTGCCTGG-CTGTTCC-3′, Glut1: Fw: 5′-CCAGC TGGGA ATCGT CG-TT-3′, Rv: 5′-CAAGT CTGCA TTGCC CATGAT-3′, Glut4: Fw: 5′-TCTGTGGGTGGCATGATCTCT-3′, Rv: 5′-GCCCT-TTTCCTTCCCAACC-3′, glucose phosphate isomerase 1: Fw: 5′-AGCGCTTCAACAACTTCAGCT-3′, Rv: 5′-CAGAATAT-GCCCATGGTTGGT-3′, phosphoglycerate mutase 1: Fw: 5′-TCCTGAAACATCTGGAAGGTATCTC-3′, Rv: 5′-CAGTG-GGCAGAGTGATGTTGAT-3′, fructose bisphosphatase 2: Fw: 5′-TGAATGCAATCCTGTGGCC-3′, Rv: 5′-TGGTTGCCA-TACCTCCTGCT-3′, pyruvate dehydrogenase kinase isoenzyme 1: Fw: 5′-GGACTTCTATGCGCGCTTCT-3′, Rv: 5′-CTGA-CCCGAAGTCCAGGAAC-3′, glycogen synthase 2: Fw: 5′-AG-GATCATTCAGAGGAACCGC-3′, Rv: 5′-CCAGTCCAGGA-GATCTGAGAGC-3′.

Tissue sampling for analysis of Glut1 and Glut4 protein levels

Skeletal muscles were homogenized in ice-cold buffer containing 250 mmol/L sucrose, 20 mmol/L 2-[4-(2-hydroxyethyl)-1-piperadinyl] ethonsulforic acid (HEPES) (pH 7.4), and 1 mmol/L EDTA, and centrifuged at 1200 g for 5 minutes. The supernatant was centrifuged at 200 000 g for 60 minutes at 4°C [13]. The resulting pellet was solubilized in Laemmli sample buffer containing dithiothreitol. Samples were subjected to Western blotting as described [14]. Antibodies used were those against Glut1 (#07-1401, Millipore, Temecula, CA), and Glut4 (SC-1606, Santa Cruz Biotechnology Inc., Santa Cruz, CA).

Muscle incubation and glucose transport

Glucose transport was measured as described [15]. Mice were fasted overnight and killed. The extensor digitorum longus (EDL) muscles were rapidly removed, each of which was mounted on the incubation apparatus and preincubated in Krebs-Ringer bicarbonate (KRB) buffer containing 2 mmol/l pyruvate for 30 min. The muscles were then incubated in KRB buffer in the absence or presence of 50 mU/ml insulin for 10 min. The buffers were kept at 37°C throughout the experiment and gassed continuously with 95% O_2 and 5% CO_2. Immediately after incubation, the muscle was transferred to KRB buffer containing 1 mmol/l 2-[^{3}H]-deoxy-d-glucose (1.5 μCi/ml) and 7 mmol/l d-[^{14}C]-mannitol (0.3 μCi/ml).

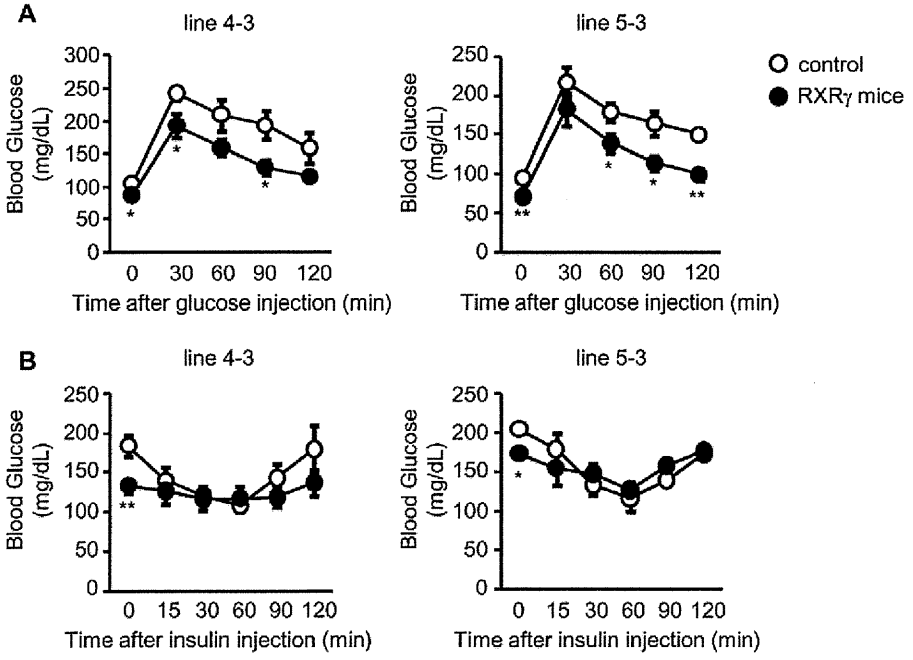

Figure 1. Glucose tolerance and insulin tolerance tests on RXRγ mice. (*A, B*) In A and B, male mice, 5 months of age, were used. The number of animals used was 6 for both control (open circles) and RXRγ (filled circles) mice of line 4-3, and 5 for both control (open circles) and RXRγ (filled circles) mice of line 5-3. * $P<0.05$ and ** $P<0.01$ compared with respective control.
doi:10.1371/journal.pone.0020467.g001

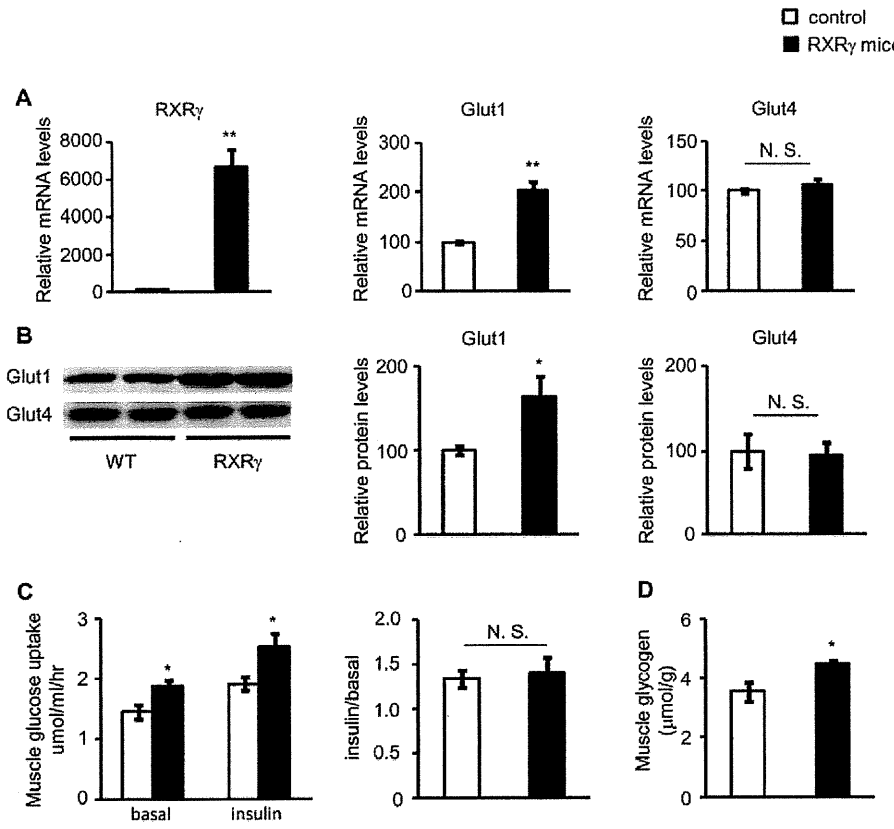

Figure 2. Levels of Glut 1 and Glut4, and glucose uptake in the skeletal muscle of RXRγ mice. (*A*) Gene expressions of *RXRγ*, *Glut1*, and *4* were examined by quantitative real-time PCR. The value for wild-type (littermates of line 4-3) mice was set at 100, and relative values are shown. (*B*) Protein levels of Glut1 and Glut4 were examined by Western blotting. Results of relative densitometric signal for Glut1 and 4 are shown. (*C*) Glucose uptake in the absence or presence of insulin and (*D*) glycogen content were increased in the skeletal muscle of RXRγ mice. Ratio of enhanced glucose uptake in the presence of insulin (insulin/basal) was similar in control and RXRγ mice. In *A, B* and *D*, the same samples were used. Mice were males of 12 weeks of age. The number of animals was 6 for both control (open bars) and RXRγ (filled bars) mice. These samples were also used in Table 1. In *C*, mice were males of 24–27 weeks of age. The number of animals was 6 for both control (open bars) and RXRγ (filled bars) mice. * $P<0.05$ and ** $P<0.01$ compared with respective control. N. S., not significant.
doi:10.1371/journal.pone.0020467.g002

Measurement of skeletal muscle glycogen

The skeletal muscle glycogen content was measured as glycosyl units after acid hydrolysis [16]. The skeletal muscle samples were minced, 50 mg of which was added to 1 ml of 0.3 M percholic acid and homogenized on ice. One ml of 1 N HCl was added and incubated for 2 h at 100°C, thereafter, 1 ml of 1 N NaOH was added at room temperature. Glucose content was examined using the F-kit glucose (Roche Diagnostics, Mannheim, Germany).

Plasmids

The coding region of mouse RXRγ and PPARδ cDNA were subcloned into a mammalian expression plasmid, pCMX [9]. The 1.5-kb 5′-flanking region of the mouse *Glut1* promoter was obtained by PCR with mouse genomic DNA (RefSeqs number: NC_000070). The PCR primers used were: Fw: 5′-GTGGTG-CGCGCCTGTAGTCC-3′ and Rv: 5′-GGCGCACTCCACG-GATGCCG-3′. The fragment was subcloned into the pGL3-basic luciferase vector (Promega Corporation, Madison, WI). The promoter regions used were: -1500 to +75, -647 to +75, and -152 to +75, counting the transcription start site as +1.

Electroporation and *in vivo* luciferase reporter analysis

In vivo electroporation was performed according to the modified method of Aihara and Miyazaki [17]. Under the pentobarbital anesthesia (30 mg/kg), bilateral quadriceps muscles from C57BL6

Table 3. Body weight, blood glucose and plasma insulin levels in RXRγ mice in clamp study.

		Wild-type	RXRγ
	Body weight (g)	24.1±1.2	23.4±0.3
Basal period	Glucose (mg/dL)	141.8±15.3	106.6±5.5*
	Insulin (ng/mL)	0.55±0.08	0.44±0.07
Clamp period	Glucose (mg/dL)	92.5±3.8	96.7±3.6
	Insulin (ng/mL)	2.60±0.20	2.38±0.33

Mice were males 13–16 weeks of age. The number of animals used was 6 for wild-type control mice and 7 for RXRγ mice.
* $P<0.05$, compared with wild-type control.
Values are the means ± SE. These mice were also used in Fig 3.
doi:10.1371/journal.pone.0020467.t003

mice (male, 12 weeks of age) were injected with 80 μg of plasmid DNA (25 μl) by using a 29-gauge needle attached to a 0.5-ml insulin syringe (Terumo Corporation, Tokyo, Japan). Square-wave electrical pulses (160 V/cm) were applied six times with an electrical pulse generator (CUY21EDIT, Nepa Gene Co. Ltd., Chiba, Japan) at a rate of one pulse per second, with each pulse being 20 ms. in duration. The electrodes were a pair of stainless steel needles inserted into the quadriceps muscles and fixed 5 mm apart. Seven days after gene delivery, the muscles were removed and subjected to analysis. Frozen muscle tissues were homogenized in ice-cold passive lysis buffer from Promega. The homogenate was centrifuged at 10,000 g for 10 min at 4°C. The supernatant was reserved for luciferase assay using Promega's dual luciferase assay kit. The luciferase activity was calculated as the ratio of firefly to Renilla (internal control) luciferase activity and represented as the average of triplicate experiments.

Computer-based DNA sequence motif search

We used MATCH software [18] (BIOBASE GmbH, Wolfenbuettel, Germany) to investigate transcriptional binding sites in the mouse *Glut1* promoter. We investigated mouse genome in the region of -1500 to +100 relative to transcription start of *Glut1* (Chr.4 11878131(+)).

cDNA microarray analysis

RNA was isolated from the skeletal muscle of sex- and age-matched RXRγ mice (line 4-3) and non-transgenic control mice

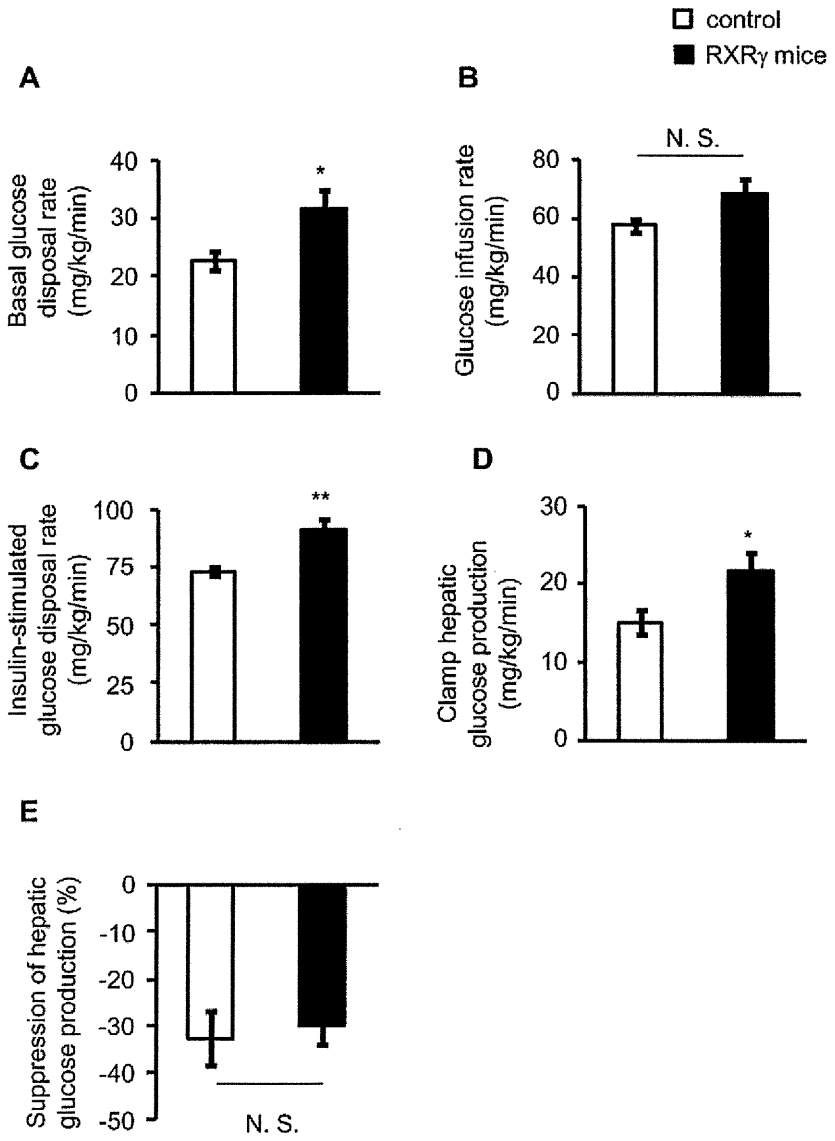

Figure 3. Hyperinsulinemic-euglycemic clamp test in RXRγ mice, fed a chow diet. (*A*) Basal glucose disposal rate, (*B*) glucose infusion rate needed to maintain euglycemia, (*C*) insulin-stimulated glucose disposal rate and (*D*) clamp hepatic glucose production (hepatic glucose production during the clamp period) (*E*) suppression of hepatic glucose production during the clamp period in RXRγ mice. Male mice, 13~16 weeks of age, were used. The number of animals used was 6 for control mice (open bars) and 7 for RXRγ mice (filled bars). * $P<0.05$ and ** $P<0.01$ compared with respective control. N. S., not significant.
doi:10.1371/journal.pone.0020467.g003

(females at 4 months of age, five samples from each group were combined). Each of the combined samples was hybridized to the Affymetrix MG430 microarray, which contains 45,102 genes, including expressed sequence tags (ESTs), and analyzed with the software Affymetrix Gene Chip 3.1. Of the 45,102 genes including ESTs analyzed, 8,054 (non-transgenic control mice) and 8,083 (transgenic) were expressed at a substantial level (absolute call is present and average difference is above 150). In order of fold changes in gene expression levels in skeletal muscle from RXRγ mice relative to control mice, genes whose expression was increased more than $2^{0.4}$-fold in RXRγ mice were listed (Dataset S1). Fold changes were calculated as an indication of the relative change of each transcript represented on the probe array. Differentially expressed genes were identified using the following criteria; 'absolute call' is present, and 'average difference' was above 250. 'Absolute call', which was calculated with this software using several markers, is an indicator of the presence or absence of each gene transcript. The 'average difference' value is a marker of the abundance of each gene, obtained by comparing the intensity of hybridization to 20 sets of perfectly matched 25-mer oligonucleotides relative to 20 sets of mismatched oligonucleotides using Affymetrix Gene Chip 3.1 software. All data of microarray is MIAME compliant and that the raw data has been deposited in a MIAME compliant database (GEO), whose accession number is GSE28448.

Gene Ontology Analysis

We used DAVID v6.7 [19] for gene ontology (GO) analysis. DAVID is a web application providing a comprehensive set of functional annotation tools to understand the biological meaning behind a large list of genes. Functional Annotation Clustering of DAVID was applied to the genes whose gene expression increased in RXRγ mice. Our GO analysis produced 62 GO terms from gene sets from RXRγ mice with increased expression compared to the wild-type mice, under the condition of $P<0.05$ (P value from Fisher's Exact Test). The obtained GO terms contained many similar functional concepts. In order to group similar GO terms, we applied the Functional Annotation Clustering tool provided by DAVID [19,20]. Twenty-three clusters were produced with genes showing increased expression. We showed one GO term of the lowest P value of all the members of an individual cluster.

Transcriptional factor binding sites analysis

We used MATCH software [18] with BKL TRANSFAC 2010.3 Release (BIOBASE GmbH, Worfenbuettel, Germany) to investigate transcriptional binding sites in the promoter regions of genes. The F-Match algorithm compares the number of sites found in a query sequence set against the background set. It is assumed, if a certain transcription factor (or factor family), alone or as a part of a *cis*-regulatory module, plays a significant role in the regulation of the considered set of promoters, then the frequency of the corresponding sites found in these sequences should be significantly higher than expected by random chance. We investigated the mouse genome in the region of -1000 to +100 relative to the transcription start of an individual gene. Statistical hypothesis testing was evaluated against housekeeping genes of mice. We investigated promoter regions of 15 genes: that is 14 'glucose metabolic process' genes with increased expression in RXRγ mice, plus Glut1. In the GO analysis, Glut1 was categorized as a 'transporter' not 'glucose metabolic process' gene.

Statistical analysis

Statistical analysis was performed using the Student's t test and analysis of variance (ANOVA) followed by Scheffe's test. Data were expressed as the mean ± SE. $P<0.05$ was considered statistically significant.

Results

Increased glucose metabolism in RXRγ mice

In our previous study, we established two lines of RXRγ mice (named lines 4-3 and 5-3) with similar expression levels of the *RXRγ* transgene and protein specifically in the skeletal muscle [9]. There was no significant difference in body weight, adipose tissue,

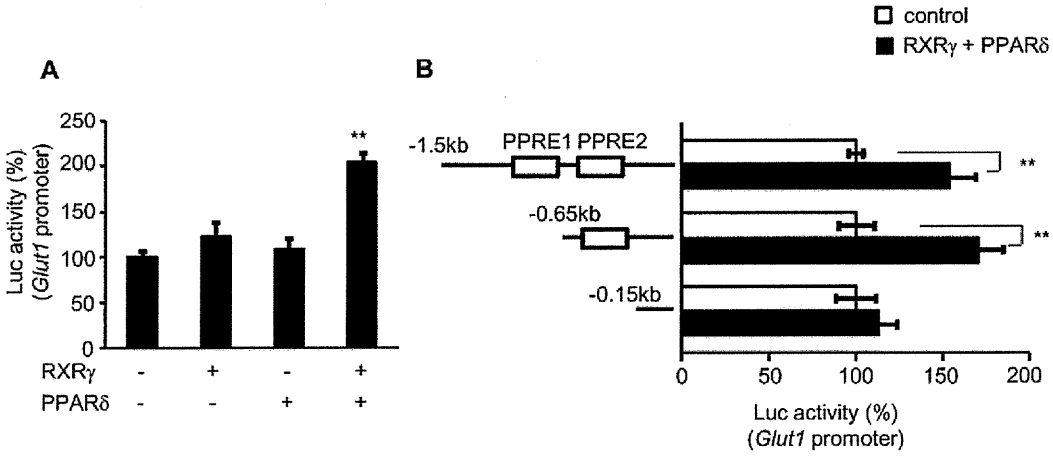

Figure 4. Transient transfection-reporter assay of the effect of RXRγ on *Glut1* promoter. (*A*) *Glut1*-Luc plasmid, with or without RXRγ and/or PPARδ expression vectors, was transfected into the quadriceps muscle of C57BL6 mice. Activation of the luciferase reporter gene was measured in relative light units and normalized to dual luciferase activity. Mean values from experiments (n = 5) are shown as fold induction, where the activity in the absence of RXRγ is the reference value (set at 100). (*B*) Schematic representations of serial deletion of *Glut1* promoter constructs are shown in the figure. Squares denote the putative PPAR/RXR binding sites. Open bars; *Glut1*-Luc without RXRγ and PPARδ expression vectors, and filled bars; *Glut1*-Luc with RXRγ and PPARδ expression vectors. The activity in the absence of RXRγ and PPARδ in each experiment for different *Glut1*-Luc construct in the reference value (set at 100). ** $P<0.01$, compared with the value of wild-type promoter in the absence of RXRγ/PPARδ.
doi:10.1371/journal.pone.0020467.g004

and liver weight between RXRγ and control mice in both lines 4-3 and 5-3, although the skeletal muscle weight was slightly lower in RXRγ mice than in control mice (Tables 1 and 2). In this study, blood glucose in RXRγ mice was lower than in control mice (fasting state, $P<0.01$ in lines 4-3 and 5-3; basal state, $P<0.05$ in line 4-3 and $P=0.06$ in line 5-3), which is consistent with our previous report [9].

To elucidate the role of the skeletal muscle RXRγ in systemic glucose metabolism, we performed glucose and insulin tolerance tests in RXRγ mice. The glucose tolerance test revealed increased glucose disposal in RXRγ mice relative to control mice (Fig. 1*A*). On the other hand, there was no significant difference in the insulin-induced hypoglycemic response between genotypes (Fig. 1*B*). These observations suggest that RXRγ mice have a higher capacity for glucose disposal with no change in insulin sensitivity.

Increased Glut1 expression and glucose uptake in the skeletal muscle of RXRγ mice

As both lines of RXRγ mice showed increased glucose metabolism, for the following experiments, we only utilized the line 4-3 of RXRγ mice (hereafter just RXRγ mice). Interestingly, mRNA expression of *Glut1* was increased in the skeletal muscle from RXRγ mice relative to control mice ($P<0.01$), whereas that of *Glut4* was unchanged (Fig. 2*A*). We also observed that Glut1 is significantly increased in RXRγ mice at the protein level ($P<0.05$), with no significant difference in Glut4 between genotypes (Fig. 2*B*). Consistently, we also found increased glucose uptake in the skeletal muscle from RXRγ mice relative to control mice ($P<0.05$), which was not further enhanced in the presence of insulin (Fig. 2*C*). We also observed increased glucose glycogen content in the skeletal muscle from RXRγ mice relative to control mice ($P<0.05$, Fig. 2*D*).

Table 4. Gene Ontology Analysis.

GO ID	GO Term	P value
GO:0060537	muscle tissue development	6.91E-05
GO:0006461	protein complex assembly	6.78E-05
GO:0008104	protein localization	1.18E-04
GO:0051146	striated muscle cell differentiation	6.06E-04
GO:0030334	regulation of cell migration	8.04E-04
GO:0006886	intracellular protein transport	0.002367379
GO:0019220	regulation of phosphate metabolic process	0.00312153
GO:0048514	blood vessel morphogenesis	0.00172443
GO:0045859	regulation of protein kinase activity	0.00563927
GO:0006915	apoptosis	0.007513423
GO:0006006	glucose metabolic process	0.002958786
GO:0006917	induction of apoptosis	0.005257921
GO:0042981	regulation of apoptosis	0.014518039
GO:0043066	negative regulation of apoptosis	0.022995692
GO:0006633	fatty acid biosynthetic process	0.012664127
GO:0032956	regulation of actin cytoskeleton organization	0.027565775
GO:0043388	positive regulation of DNA binding	0.021966247
GO:0006469	negative regulation of protein kinase activity	0.045894089
GO:0016477	cell migration	0.045248711
GO:0030521	androgen receptor signaling pathway	0.027402985
GO:0055003	cardiac myofibril assembly	0.020081231
GO:0000165	MAPKKK cascade	0.032045331
GO:0046825	regulation of protein export from nucleus	0.027402985

738 genes up-regulated in RXRγ mice compared with wild-type mice by microarray (Listed in Dataset S1) were classified into GO functional annotations, as described in Methods.
doi:10.1371/journal.pone.0020467.t004

Increased basal glucose disposal rate of RXRγ mice

To gain further insight into the glucose metabolism in RXRγ mice, we performed a hyperinsulinemic-euglycemic clamp study. Plasma insulin concentrations during the basal period were similar between genotypes (Table 3). The basal glucose disposal rate was significantly increased in RXRγ mice than in the controls ($P<0.05$, Fig. 3*A*), supporting that an insulin-independent increase in glucose uptake occurred, as observed in Fig. 2*C*. Meanwhile, we observed that the rate of glucose infusion needed to maintain euglycemia (glucose infusion rate) was similar between genotypes (Fig. 3*B*), which is consistent with the result of the insulin tolerance test (Fig. 1*B*). On the other hand, insulin-stimulated glucose disposal rate was higher in RXRγ mice than in the controls ($P<0.01$, Fig. 3*C*), which probably reflects the increased basal glucose disposal rate (Fig. 3*A*). Also, the clamp hepatic glucose production (hepatic glucose production during the clamp period) was higher in RXRγ mice than in the controls ($P<0.05$, Fig. 3*D*). Meanwhile, hepatic glucose production was similarly suppressed by insulin in both genotypes (Fig. 3*E*). Together, these data support the idea that Glut1, an insulin-independent glucose transporter, is involved in the increased glucose disposal in the skeletal muscle of RXRγ mice.

Activation of the *Glut1* promoter by combination of RXRγ and PPARδ in the skeletal muscle *in vivo*

To examine whether RXRγ directly activates mRNA expression of *Glut1*, we performed the *in vivo* luciferase reporter analysis

Table 5. List of 'glucose metabolic process' genes.

Gene Symbol	Gene description
Atf3	Activating transcription factor 3
Bpgm	2-3-bisphosphoglycerate mutase
Dcxr	Dicarbonyl L-xylulose reductase
Fbp2	Fructose bisphosphatase 2
Gbe1	Glucan branching enzyme 1
Gpd1l	Glycerol-3-phosphate dehydrogenase 1-like
Gpi1	Glucose phosphate isomerase 1
Gys2	Glycogen synthase 2
Igf2	Insulin-like growth factor 2
Mat2b	Methionine adenosyltransferase II, beta
Nisch	Nischarin; an imidazoline receptor
Pdk1	Pyruvate dehydrogenase kinase isoenzyme 1
Pgam1	Phosphoglycerate mutase 1
Pgd	Phosphogluconate dehydrogenase
Pgm2l1	Phosphoglucomutase 2-like 1
Phkb	Phosphorylase kinase beta

Up-regulated genes in RXRγ mice in the microarray, classified as 'glucose metabolic process' genes by GO analysis, as described in Method. Genes are listed in alphabetic order of gene symbol. Glut1, which appeared in the up-regulated list in the microarray (Dataset S1), is not included in this list, as it was classified 'transporter' in the GO analysis.
doi:10.1371/journal.pone.0020467.t005

using *Glut1* promoter (*Glut1*-Luc). In this study, the activity of *Glut1*-Luc was marginally enhanced by RXRγ alone (Fig. 4*A*). Moreover, we found no significant activation of the *Glut1*-Luc by PPARδ, an important regulator of glucose as well as lipid metabolism in the skeletal muscle [7]. Interestingly, combination of RXRγ and PPARδ resulted in significant increase in *Glut1*-Luc activity in the skeletal muscle *in vivo* ($P<0.01$) (Fig. 4*A*). Motif search analysis revealed two putative PPAR-responsive elements (PPRE1 and PPRE2) (−657/−645 and −520/−508, respectively) in the mouse *Glut1* promoter. A series of deletion mutant analysis showed that combination of RXRγ and PPARδ activates the regions of −1500/+75, and −647/+75, but not the region of −152/+75 in the *Glut1* promoter (Fig. 4*B*), suggesting that the second putative PPRE (−520/−508) is involved in the RXRγ/PPARδ-induced *Glut1* promoter activation.

Microarray and bioinformatics analyses of up-regulated gene in RXRγ mice

In order to gain insight into the gene expression change in RXRγ mice, we performed microarray analysis. As shown in Dataset S1, 738 genes were up-regulated in the analysis. As expected, Glut1 expression was increased in the microarray data. Also, SREBP1c expression, which we previously reported [9], was increased in the microarray data. Using the data, we performed GO analysis to determine if genes, up-regulated in RXRγ mice, are associated with particular biological processes. Our GO analysis revealed genes with increased expression in the RXRγ mice in various categories (Table 4), including 'glucose metabolic process' genes and 'fatty acid biosynthetic process' genes, which indicated that overexpression of RXRγ affects the expression of many genes. The up-regulated genes categorized as glucose metabolism genes in GO term are listed in Table 5. Among them, we confirmed enhanced gene expression by quantitative real time PCR (Fig. 5), supporting the microarray data reliable. Moreover, we calculated the ratio of putative transcription factor binding motifs in glucose metabolism genes, which were up-regulated in RXRγ mice. In the sample, several motifs showed statistical significance (Table 6) ($P<0.05$), including PPAR responsive elements. These data suggest that glucose metabolism genes up-regulated in RXRγ mice are possible target genes of the RXRγ and PPAR heterodimer.

Discussion

RXRγ is a nuclear receptor-type transcription factor that is expressed abundantly in the skeletal muscle and is regulated by nutritional conditions. Treatment of obese and diabetic mice with RXR pan-agonists (agonists for all the RXR isoforms) has improved glucose metabolism in mice [21–23], suggesting the beneficial effect of RXR on diabetes. However, which RXR isoform(s) are involved and where they work to improve diabetes

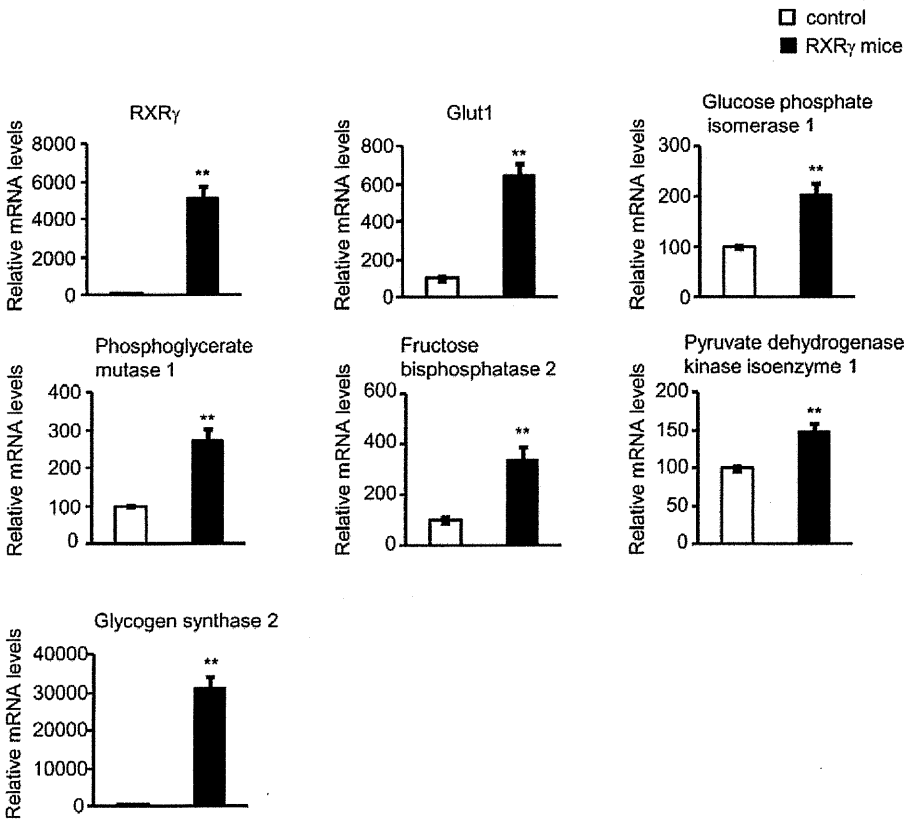

Figure 5. Levels of 'glucose metabolic process' gene expression in the skeletal muscle of RXRγ mice. Representative gene expressions of 'glucose metabolic process genes' analyzed by microarray and GO analysis (Table 5) were examined by quantitative real-time PCR. The value for wild-type (littermates of line 4-3) mice was set at 100, and relative values are shown. Mice were females of 4 months of age. The number of animals was 6 for both control (open bars) and RXRγ (filled bars) mice. These samples were also used in microarray analysis (Dataset S1). * $P<0.05$ and ** $P<0.01$ compared with respective control. N. S., not significant.
doi:10.1371/journal.pone.0020467.g005