To collect large amounts of cancer genome data from different ancestry groups and epidemiological backgrounds, we currently need to combine data from multiple institutes that apply individual analytical platforms. An important caveat in multicenter trans-ancestry analysis has been the possibility that ancestry-specific signatures can be biased by experimental or analytical differences. To avoid this potential bias, we processed the DNA from 99 Japanese HCC cases using the sequencing and analysis pipeline at the United States–based Baylor College of Medicine. Using this data set from a single center, we replicated exactly the same signatures in each population (**Supplementary Fig. 26**). We also examined the distribution of signatures among three centers using Japanese male samples and confirmed that similar distributions were seen among the three centers (**Supplementary Fig. 27**). Furthermore, we analyzed whole-genome sequencing data for 88 Chinese HCC samples[19] and successfully identified HCC signatures B and C in this independent data set (**Supplementary Fig. 28**).

Outcome analysis from mutational signatures

We analyzed the derived NMF signatures to determine whether any signature or signature component was associated with differences in outcome in the HCC cohort. NMF signature values were merged with annotated clinical data. We performed calculations using standardized signature values to control for differences in the mutation rate between the subjects. Multivariate analysis with the Cox proportional hazards model (**Supplementary Fig. 29** and **Supplementary Tables 22–26**) indicated that histological grade, HCC signature B and the interaction with HCC signature A (but not with HCC signature C) were significant predictors of outcome.

DISCUSSION

The present trans-ancestry liver cancer genome study first identified mutational signatures that are independent of hepatitis virus infection and contribute more to the Asian cases than to ones of European ancestry (**Supplementary Tables 27**). One signature, characterized by AT>AC mutations, was predominant in Japanese males, whereas the other, featuring CTG>CAG mutations, was found more frequently in tumors from Asians living in the United States. These correlations may highlight deeper intra-ancestry diversity and/or environmental contributions, and sex bias might further affect downstream target genes and molecular features in HCC[36]. As several genetic loci are associated with individual HCC risk together with HBV and/or HCV infection[37,38], somatic and germline genome interaction might also be important to consider. Notably, these signatures were not evident in IHCC for Japanese cases (data not shown), suggesting that they are unique properties of HCC. The causes of these signatures remain unknown, but skewed transcriptional strand biases in characteristic sequence contexts strongly imply the presence of specific, previously unexplored mutational processes, which profoundly influence tumor genome constitution and behavior.

With 503 cases, this study is the largest liver cancer genome analysis thus far, enabling the formation of a more thorough picture of the mutational landscape of HCC than ever before. In addition to identifying a large number of significantly mutated genes, we have also identified recurrent alterations of 9 of the 14 core genes making up the SWI/SNF complex. We also find a combination of hotspot *TERT* promoter and *ATRX* mutations, along with focal amplification and virus genome integration in the *TERT* locus, in more than 68% of HCC cases regardless of virus subtype. These findings show that *TERT* is a central driver gene and a promising molecular target[39] in HCC. The targeting of high-prevalence mTOR-PIK3CA pathway activation and antiproliferative activity in HCC cells by chemical inhibition should also offer new therapeutic opportunities. In addition, newly identified alterations in the chromatin-remodeling complex and metabolic enzymes are expected to be associated with cancer-specific epigenetic and metabolomic features.

URLs. DNAcopy, http://www.bioconductor.org/packages/2.13/bioc/html/DNAcopy.html; R software, http://www.R-project.org/; R survival package, http://CRAN.R-project.org/package=survival/; HGSC Mercury analysis pipeline, https://www.hgsc.bcm.edu/software/mercury; GRCh38 human reference genome, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/; BWA2, http://bio-bwa.sourceforge.net/; GATK4, http://www.broadinstitute.org/gatk/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000389, the ICGC database (http://www.icgc.org/) and the database of Genotypes and Phenotypes (dbGaP) under accession phs000509.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This study was supported by Grants-in-Aid from the Ministry of Health, Labour and Welfare of Japan for the third-term Comprehensive 10-Year Strategy for Cancer Control, grants from the US National Human Genome Research Institute (NHGRI; 5U54HG003273) and National Cancer Institute (NCI; HHSN261201000053C and P30 CA125123), the Program for Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation (NIBIO, Japan) and the National Cancer Center Research and Development Funds (23-A-8, Japan). The National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan. The supercomputing resource SHIROKANE was provided by the Human Genome Center at the University of Tokyo (http://sc.hgc.jp/shirokane.html).

AUTHOR CONTRIBUTIONS

Study design: Y.T., K.T., K.R.C., H.U., M.K., D.A.W., H.A. and T.S. Sequencing data generation: K.T., D.M.M., F.H., H. Doddapaneni, H. Dinh, Y.A., K.G., K.W., M.-C.G., T.U., S.O., N.O., M.W. and Y.Z. Data analysis: Y.T., K.T., K.R.C., H.U., M.K., S.T., L.A.D., B.L.S., E.S., S.Y., H.N., M.L., N.H., K.W., K.G., M.D., G.N., D.A.W. and T.S. Statistical analysis: Y.T., K.R.C., H.U., K.T., C.J.C., M.K., S.T. and S.Y. Molecular analysis: Y.A. and T.S. Sample acquisition and clinical data collection: M.-C.G., K.S., Y.M., J.A.G., H.O., A.H., J.S., R.C., J.G., S.I., M.T., T.O., N.K., T.K., T.T. and M.F. Manuscript writing: Y.T., K.T., K.R.C., H.U., C.J.C., L.A.D., B.L.S., M.K., D.A.W., H.A. and T.S. Project oversight: D.A.W., R.A.G., H.A. and T.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).
2. Forner, A., Llovet, J.M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **379**, 1245–1255 (2012).
3. El-Serag, H.B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264–1273 (2012).
4. Yu, J., Shen, J., Sun, T.T., Zhang, X. & Wong, N. Obesity, insulin resistance, NASH and hepatocellular carcinoma. *Semin. Cancer Biol.* **23**, 483–491 (2013).
5. Augustine, M.M. & Fong, Y. Epidemiology and risk factors of biliary tract and primary liver tumors. *Surg. Oncol. Clin. N. Am.* **23**, 171–188 (2014).
6. Tanaka, K., Sakai, H., Hashizume, M. & Hirohata, T. Serum testosterone:estradiol ratio and the development of hepatocellular carcinoma among male cirrhotic patients. *Cancer Res.* **60**, 5106–5110 (2000).
7. International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).

© 2014 Nature America, Inc. All rights reserved.

8. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
9. Wang, K. *et al.* Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* **58**, 706–717 (2013).
10. Sung, W.K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
11. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
12. Killela, P.J. *et al. TERT* promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. USA* **110**, 6021–6026 (2013).
13. Nault, J.C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
14. Li, Y. & Tergaonkar, V. Noncanonical functions of telomerase: implications in telomerase-targeted cancer therapies. *Cancer Res.* **74**, 1639–1644 (2014).
15. Heaphy, C.M. *et al.* Altered telomeres in tumors with *ATRX* and *DAXX* mutations. *Science* **333**, 425 (2011).
16. Hoffmeyer, K. *et al.* Wnt/β-catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549–1554 (2012).
17. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
18. Li, M. *et al.* Inactivating mutations of the chromatin remodeling gene *ARID2* in hepatocellular carcinoma. *Nat. Genet.* **43**, 828–829 (2011).
19. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
20. Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).
21. Tetsu, O. & McCormick, F. β-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
22. Pai, R. *et al.* Inhibition of fibroblast growth factor 19 reduces tumor growth by modulating β-catenin signaling. *Cancer Res.* **68**, 5086–5095 (2008).
23. Motohashi, H. & Yamamoto, M. Nrf2-Keap1 defines a physiologically important stress response mechanism. *Trends Mol. Med.* **10**, 549–557 (2004).
24. Zhang, D.D., Lo, S.C., Cross, J.V., Templeton, D.J. & Hannink, M. Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. *Mol. Cell. Biol.* **24**, 10941–10953 (2004).
25. Song, L.N. & Gelmann, E.P. Silencing mediator for retinoid and thyroid hormone receptor and nuclear receptor corepressor attenuate transcriptional activation by the β-catenin–TCF4 complex. *J. Biol. Chem.* **283**, 25988–25999 (2008).
26. Eissenberg, J.C., Wong, M. & Chrivia, J.C. Human SRCAP and *Drosophila melanogaster* DOM are homologs that function in the Notch signaling pathway. *Mol. Cell. Biol.* **25**, 6559–6569 (2005).
27. Monroy, M.A. *et al.* SNF2-related CBP activator protein (SRCAP) functions as a coactivator of steroid receptor–mediated transcription through synergistic interactions with CARM-1 and GRIP-1. *Mol. Endocrinol.* **17**, 2519–2528 (2003).
28. Hood, R.L. *et al.* Mutations in *SRCAP*, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome. *Am. J. Hum. Genet.* **90**, 308–313 (2012).
29. Nelson, R.A. *et al.* Floating-Harbor syndrome and intramedullary spinal cord ganglioglioma: case report and observations from the literature. *Am. J. Med. Genet. A.* **149A**, 2265–2269 (2009).
30. Iyer, G. *et al.* Genome sequencing identifies a basis for everolimus sensitivity. *Science* **338**, 221 (2012).
31. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
32. Pleasance, E.D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
33. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
34. Poon, S.L. *et al.* Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci. Transl. Med.* **5**, 197ra101 (2013).
35. Goedde, H.W. *et al.* Population genetic studies on aldehyde dehydrogenase isozyme deficiency and alcohol sensitivity. *Am. J. Hum. Genet.* **35**, 769–772 (1983).
36. Keng, V.W. *et al.* Sex bias occurrence of hepatocellular carcinoma in Poly7 molecular subclass is associated with *EGFR*. *Hepatology* **57**, 120–130 (2013).
37. Zhang, H. *et al.* Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat. Genet.* **42**, 755–758 (2010).
38. Kumar, V. *et al.* Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat. Genet.* **43**, 455–458 (2011).
39. Harley, C.B. Telomerase and cancer therapeutics. *Nat. Rev. Cancer* **8**, 167–179 (2008).

© 2014 Nature America, Inc. All rights reserved.

npg

ONLINE METHODS

DNA preparation, DNA capture and sequencing. The tissues and clinical information used in this study were obtained under informed consent and approval of the institutional review boards of each institute. DNA was extracted from liver cancer tissue and matched non-cancerous liver tissues or blood using a general protocol for genome sequencing. Exome capture was carried out using the SureSelect Human All Exon V3 or V4 plus kit depending on the samples (**Supplementary Table 28**). Preparation of sequencing libraries, DNA capture methods and Illumina sequencing were carried out as described in the **Supplementary Note**.

Mutation calling. *Mutation calling (National Cancer Center Research Institute).* Paired-end reads were aligned to the human reference genome (GRCh37) using the Burrows-Wheeler Aligner (BWA)[40] for both tumor and normal samples. Probable PCR duplications, for which paired-end reads aligned to the same genomic position, were removed, and pileup files were generated using SAMtools[41] and a program developed in house. Details on our filtering conditions are provided in **Supplementary Tables 29** and **30**.

Mutation calling (Research Center for Advanced Science and Technology). Next-generation sequencing reads were mapped to the human genome (hg19) using BWA and Novoalign independently. Reads with a minimal editing distance to the reference genome were taken to represent optimal alignments. Then, bam files were locally realigned with SRMA. Normal-tumor pair bam files were processed using an in-house genotyper (karkinos), with the variants further filtered to remove all variants observed fewer than four times or present at an allele frequency of less than 0.12 after adjustment for tumor sample purity. The variants also had to have a score of greater than Q20 (representing the root mean square of mapping quality). In addition, reads harboring the variant had to be observed in both forward and reverse orientation. If a variant was present in reads of only one orientation, we checked for strand bias using a *t* test comparing these reads to the reads without the variant, and variants with a *P* value of <0.03 for strand bias were rejected. Variants also had to be called in different sequence cycles and have at least one call that was outside of 3% of read ends. Variants could not be located within 5 bp of an indel call, and variants where the mean base quality of the supporting reads was lower than 10 on the Phred scale were removed. Germline variants having an allelic frequency of greater than 0.1 were collected for 50 normal liver exome samples and used as the panel of normal variants. Any variant that was observed in this panel with a population frequency of greater than 2% was filtered out. Finally, variants also observed in the paired normal sample with an allelic frequency of greater than 3% and sites registered in dbSNP Build 134 with validated status were removed.

Mutation calling (Baylor College of Medicine). Initial sequence analysis was performed using the Human Genome Sequencing Center (HGSC) Mercury analysis pipeline. First, the primary analysis software on the instrument produced bcl files that were transferred off the instrument to the HGSC analysis infrastructure by the HiSeq Real-Time Analysis module. Once each run was complete and all bcl files were transferred, Mercury ran the vendor's primary analysis software (CASAVA), which demultiplexed pooled samples and generated sequence reads and base call confidence values (qualities). In the next step, reads were mapped to the GRCh37 human reference genome using BWA (BWA2), producing a bam3 (binary alignment/map) file. The third step involved quality recalibration (using GATK4) and, where necessary, the merging of bam files for separate sequence events into a single sample-level bam file. Sorting of bam files, duplicate read marking and realignment to improve indel discovery all occurred at this step.

Processing the significantly mutated genes. The significantly mutated genes for this study were identified through three separate tests as described below (an aggregated somatic alteration method, MutSigCV[42] and an inactivation bias method), and the resulting gene lists were combined in a final table of significantly mutated genes (**Supplementary Table 13**).We also developed two tests to detect bias in the mutation list that could be a source of artifact (K.R.C., E.S., L.A.D. and D.A.W., unpublished data). One of these tests examined sequencing center bias, and the other examined bias in mutation allele fraction, which if consistently low would suggest that a gene was a passenger rather than a driver. Genes in the final combined table that failed these bias tests were removed from the final list of significantly mutated genes. Data from each process are shown in **Supplementary Tables 7–12**, and the steps are shown schematically in **Supplementary Figure 16**.

Aggregated somatic alteration method. We identified significantly altered genes by aggregating somatic substitutions, short indels, homozygous deletions and focal amplifications. We initially estimated the expected number of each alteration in each gene as follows.

First, the substitution rate was estimated by dividing the number of synonymous mutations in a sample by the number of synonymous sites in the genome. For each gene, the expected number of substitutions was calculated by multiplying the substitution rate by the number of nonsynonymous sites and splice sites in the gene. Because the substitution rate at CpG sites was much higher than that in other regions, the substitution rates and expected numbers of substitutions at CpG and non-CpG sites were estimated separately using the following equation:

$$\mathrm{EN} = \sum_{i=1}^{n}\left(\frac{M_{\mathrm{CG}_i} \times N_{\mathrm{CG}}}{S_{\mathrm{CG}} \times C_i} + \frac{M_{\mathrm{NCG}_i} \times N_{\mathrm{NCG}}}{S_{\mathrm{NCG}} \times C_i}\right)$$

where n is the number of samples, M_{CG_i} is the number of synonymous mutations at CpG sites in the ith sample, M_{NCG_i} is the number of synonymous mutations in non-CpG sites in the ith sample, S_{CG} is the number of synonymous sites at CpG sites in the genome, S_{NCG} is the number of synonymous sites at non-CpG sites in the genome, N_{CG} is the number of nonsynonymous sites and splice sites at CpG sites in a gene, N_{NCG} is the number of nonsynonymous sites and splice sites at non-CpG sites in a gene, C_i is the fraction of sequence coverage in the genome in the ith sample (usually the fraction of coding regions that have more than 20× sequence depth for whole-exome sequencing) and EN is the expected number of nonsynonymous and splice-site substitutions in a gene.

Second, the coding indel rate was estimated by dividing the number of coding indels in a sample by the number of coding sites in the genome. For each gene, the expected number was calculated by multiplying the coding indel rate by the coding length of a gene as follows:

$$\mathrm{EI} = \sum_{i=1}^{n}\frac{I_i \times L}{S \times C_i}$$

where I_i is the number of coding indels in the ith sample, S is the number of coding sites in the genome, L is the coding length of the gene and EI is the expected number of coding indels in a gene.

Third, as regions of focal amplification and homozygous deletion are much broader than gene regions, the number of focal amplifications and homozygous deletions affecting a gene in a sample is 0 or 1 and is not influenced by gene length. Therefore, the expected number of these events is the same for all genes. The expected numbers of focal amplifications and homozygous deletions were estimated separately by dividing the total length of the focal amplification or homozygous deletion region in a sample by the length of the genome as follows:

$$\mathrm{EA} = \sum_{i=1}^{n}\frac{A_i}{G \times C_i}$$

$$\mathrm{ED} = \sum_{i=1}^{n}\frac{D_i}{G \times C_i}$$

where A_i is the total length of focal amplifications in the ith sample, D_i is the total length of homozygous deletions in the ith sample, G is the length of the genome, EA is the expected number of focal amplifications in the gene and ED is the expected number of homozygous deletions in the gene.

Fourth, the expected number of protein-altering mutations was calculated by aggregating the expected numbers of nonsynonymous and splice-site substitutions in CpG and non-CpG sites, coding indels, focal amplifications and homozygous deletions as follows:

$$\mathrm{E} = \mathrm{EN} + \mathrm{EI} + \mathrm{EA} + \mathrm{ED}$$

where E is the expected number of protein-altering mutations in a gene.

© 2014 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3126

Fifth, tests of the significance of each gene were performed by assuming a Poisson distribution of mutation number. Adjustment for multiple testing was performed using the Benjamini-Hochberg method[8].

Inactivation bias method. The number of missense mutations was compared to the number of inactivating mutations (nonsense, frameshift and splice site) using a χ^2 test.

Analysis of sequencing center bias. Because multiple centers participated in this study, we sought to control for the influence of differences in mutation calling strategy, which might promote a gene to significance merely because of a bias in the variant callers used. Many studies do not use multiple callers and therefore have no way to control for these biases. For each gene with more than five variants, we counted the number of subjects for whom the gene was called for each center. These counts were compared to the total number of subjects using the χ^2 test. The results of the analysis for center bias are presented in **Supplementary Table 11**.

Analysis of subclone bias. Oncogenic driver events in a given tumor should exhibit allele fractions that are roughly the same as the mean allele fraction for the entire sample for any given subject. We separated oncogenic (driver) events from recurrent passenger events by comparing the allele fraction of mutations in candidate genes to the matched mean allele fraction of the sample, across all samples in the cohort. First, the mean somatic allele fraction was calculated for each subject (AF*s*). Next, for each variant in each gene, the allele fraction for the variant (AF*g*) was compared to the AF*s* in the respective subject. We calculated the fraction of events where AF*g* was less than AF*s* and generated a *P* value using a one-sided pairwise Wilcoxon test where the alternative hypothesis was that AF*g* was less than AF*s* (always with respect to the relevant subject). The histogram of all allele fraction biases (sum(AF*g* < AF*s*)/*n*, where *n* is variant count) is shown in **Supplementary Figure 30**. Selected significantly mutated genes are plotted individually to show how known drivers are distributed by this test. Note that several tumor-suppressor genes exhibited enrichment above the average allele fractions (for example, *RB1* and *TP53*). In these cases, the genes were typically both mutated and underwent loss of heterozygosity (LOH) for the wild-type allele. The results of subclone bias testing for all genes with more than five mutations are presented in **Supplementary Table 12**.

Copy number analysis, tumor purity and adjustment of mutated allele frequency. Initial copy number estimates were obtained by comparing read depth information for tumor and normal samples using VarScan2 (ref. 43). Depth estimates were then segmented using circular binary segmentation (CBS) as implemented in the DNAcopy package in R[44]. We used the JISTIC[45] program to generate a combined copy number matrix file. The VCRome2.1 probe locations were used as marker positions for copy number analysis. We then used JISTIC to calculate the significance for copy number gains and losses. Focal amplification at the *TERT* locus was determined using the average read depth of each captured target region.

Evaluation of tumor ploidy and purity. Using bam files from normal and tumor samples, read depth was calculated for each captured target region. After normalization by the number of total reads and GC content using regression analysis, the tumor/normal depth ratio was calculated, and values were smoothed using the moving average. Copy number peaks were then estimated using wavelet analysis, and each peak was approximated using Gaussian models. Hidden Markov models (HMMs) with the calculated Gaussian peaks were constructed, and copy number peaks were linked to genomic regions. The allelic imbalance for each copy number peak was calculated on the basis of heterozygous SNPs within the assigned region, and imbalance information and peak distances were further analyzed by model fitting where the optimal solution for a copy number peak was determined using vector matching, yielding estimated copy number and tumor purity and ploidy data simultaneously. Detailed algorithms will be described elsewhere (H.U., S.Y., K.T. and H.A., unpublished data).

HBV integration analysis. *HBV integration detection.* Viral genomes (HBV, NC_003977.1; HPV-16, NC_001526; HPV-18, NC_001357; HTLV-1, NC_001436) were downloaded from NCBI and included in the reference files when reads were mapped by BWA. No read was mapped to a virus other than HBV. To achieve more precise HBV mapping, we mapped all reads to the HBV reference sequence using the *q*-gram and Smith-Waterman method. An 11-mer *q*-gram was first applied to both strands of the HBV reference, and reads with 15 or more hits were subjected to Smith-Waterman alignment. The other end of each read was mapped to the hg19 human sequence using BWA. Finally, HBV integration sites were clustered by genomic position with a window size of 300 bp (approximately equal to the library fragment size), and sites with more than three supporting reads were used in the analysis.

Randomization test of HBV integration and copy number breakpoints. The 7,891 copy number breakpoints and 1,039 HBV integration sites were detected in 70 HBV-positive samples. Coexistence of the copy number breakpoints and HBV integration sites was examined using a 500-kb window size. To show statistical significance, we performed a randomization test by switching the position of the HBV integration sites to the same number of integration sites observed in the normal sample of other cases. We repeated this switching 100,000 times to yield distributions and estimated the *P* value.

Verification of single-nucleotide variation. We validated our mutation calls for frequently mutated genes (**Supplementary Table 31**) by resequencing samples using the Ion Proton sequencer (Life Technologies). Details are provided in the **Supplementary Note**.

Sanger sequencing of the *TERT* promoter. Bidirectional sequencing of the *TERT* promoter region was completed for 519 HCC samples. PCR runs were set up using 20 ng of genomic DNA, 10 μM manually designed primers (**Supplementary Table 32**) and KAPA HiFi DNA polymerase (Kapa Biosystems, KK2612). Touchdown PCR was performed with the following parameters: an initial denaturation at 98 °C for 5 min followed by 10 cycles of 98 °C for 30 s, 72 °C for 30 s and 72 °C for 1 min (decreasing the annealing temperature by 1 °C per cycle). The reaction then continued with 30 cycles of 98 °C for 30 s, 63 °C for 30 s and 72 °C for 1 min followed by a final extension at 72 °C for 5 min. The PCR products were purified with a 1:15 dilution of Exo-SAP, diluted by 0.6× and cycle sequenced for 25 cycles using a 1:64 dilution of BigDye Terminator v3.1 reaction mix (Applied Biosystems, 4337456). Finally, reactions were precipitated with ethanol, resuspended in 0.1 mM EDTA and analyzed on ABI 3730xl sequencing instruments using the Rapid36 run module and 3xx base-caller. SNPs were identified using SNP Detector software and were validated visually with Consed.

Analysis of mutation patterns and signatures. Mutation patterns for cases with hypermutation and IHCC cases were distinct from those for HCC cases (**Supplementary Figs. 4** and **21**), and cases with a small number of mutations cannot accurately represent the frequency of mutational patterns; therefore, cases with hypermutation, IHCC cases and cases with fewer than 40 mutations were excluded from further mutation pattern analysis.

The number of each of 96 possible somatic substitution types, C>A/G>T, C>G/G>C, C>T/G>A, T>A/A>T, T>C/A>G and T>G/A>C with the bases immediately 5′ and 3′ to each substitution in coding regions, was counted for each sample. The frequency of each of these substitutions was determined by dividing each count by the total number of substitutions, and the resulting frequencies were used for principal-component analysis. Principal-component analysis was implemented using the R command prcomp with the scaling option on. We used Wilks' λ test to evaluate the significance of the mean vector differences in different populations. We applied NMF to the 96-substitution pattern using published software[13], running 1,000 iterations of NMF with each NMF run iterated until convergence was achieved (10,000 iterations without change) or until the maximum number of 1,000,000 iterations was reached. We used another published software package[14] for model selection in NMF (selecting the input number of mutational signatures). Details on model selection for our NMF analysis are provided in the **Supplementary Note** and **Supplementary Figures 31–35**.

Pathway analysis. We used gene sets from MsigDB C2.all as pathway data sets. To assign *P* values representing the enrichment of mutations in pathways, we first checked whether a gene had at least one non-silent mutation or overlapped with focal CNAs for each sample in a given pathway (gene set). If so, we referred to such a gene as a 'mutated gene' for a sample. We then computed a population frequency for pathways with at least one mutated gene in the given

© 2014 Nature America, Inc. All rights reserved.

pathway and divided the frequency by the total length of the unioned exons of all genes in the pathway to correct for the greater number of mutations in longer genes. This quotient was used as a test statistic. We used a bootstrapping approach to calculate *P* values. In the bootstrapping approach, we randomly selected as many genes as in the given pathway from all genes in the genome and then calculated the statistic. We repeated this sampling 2,000 times, calculating a fraction corresponding to the number of sampling results in which a statistic value was greater than or equal to the value in the observed data. This fraction was used as a *P* value.

To find intensively mutated gene modules in liver cancer tissue using the identified significantly mutated gene sets from MsigDB analysis, we used Pathway Commons[15] data for the whole unbiased human gene network and integrated the gene sets into this network. All pairs of gene relationships were weighted by how many mutated genes were shared by the two genes (shared ratio). These gene relationships constituted the gene network. The whole network was split into one large connected network and some isolated small networks. To extract gene modules, we recursively eliminated edges with low shared ratio values and distinguished into the smaller modules. Although the recursive edge elimination procedure gradually clarifies tightly connected gene modules, gene modules were rarely isolated from the whole network. Using this compression process and some additional manual curation, we finally selected ten representative modules that were intensively mutated in liver cancer tissues. We took essentially the same approach as described above to calculate *P* values for mutation enrichment and mutual exclusivity for a gene pair or combination of modules. For mutation enrichment, we used all genes in a pair of modules. For mutual exclusivity, if a module had at least one mutated gene, we referred to such a module as an 'impaired module' and computed a frequency of impaired modules for each sample.

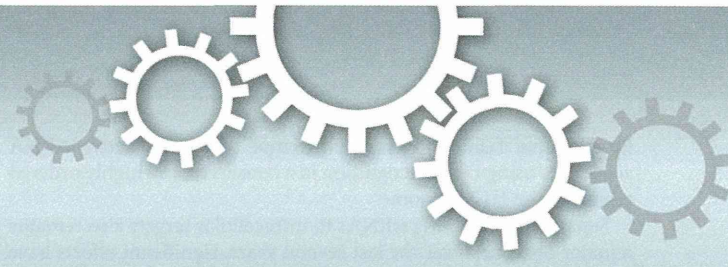
Outcome analysis from non-negative matrix factorization signatures. NMF signature values were merged with annotated clinical data for our cohort. We performed calculations using standardized signature values to control for differences in mutational rate among the subjects. For the standardized data, the contributions of each signature within a subject summed to 1. We performed Cox proportional hazards analysis[46] using the R[44] survival package, factoring in all three signature components (signature A, signature B and signature C), age at diagnosis and histological tumor grade.

40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
41. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
42. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
43. Koboldt, D.C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
44. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2010).
45. Sanchez-Garcia, F., Akavia, U.D., Mozes, E. & Pe'er, D. JISTIC: identification of significant targets in cancer. *BMC Bioinformatics* **11**, 189 (2010).
46. Cox, D.R. & Oakes, D. *Analysis of Survival Data* (Chapman & Hall/CRC, Boca Raton, FL, 1984).

© 2014 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3126

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
NUCLEIC-ACID THERAPEUTICS
HEPATITIS C VIRUS

Received
17 March 2014

Accepted
4 April 2014

Published
23 April 2014

Correspondence and requests for materials should be addressed to M.K. (kohara-mc@igakuken.or.jp)

* These authors contributed equally to this work.

In vivo therapeutic potential of Dicer-hunting siRNAs targeting infectious hepatitis C virus.

Tsunamasa Watanabe[1,2,6*], Hiroto Hatakeyama[3*], Chiho Matsuda-Yasui[1*], Yusuke Sato[3*], Masayuki Sudoh[4], Asako Takagi[1], Yuichi Hirata[1,2], Takahiro Ohtsuki[1], Masaaki Arai[5], Kazuaki Inoue[2], Hideyoshi Harashima[3] & Michinori Kohara[1]

[1]Department of Microbiology and Cell Biology, Tokyo Metropolitan Institute of Medical Science, Tokyo 156-8506, Japan, [2]Division of Gastroenterology, Showa University Fujigaoka Hospital, Yokohama, Japan, [3]Laboratory of Innovative Nanomedicine, Faculty of Pharmaceutical Sciences, Hokkaido University, Hokkaido 060-0812, Japan, [4]Kamakura Research Laboratories, Chugai Pharmaceutical Co., Ltd., Kamakura, Kanagawa 247-8530, Japan, [5]Advanced Medical Research Laboratory, Mitsubishi Tanabe Pharma Corporation, 1000, Kamoshida-cho, Aoba-ku, Yokohama 227-0033, Japan, [6]Present address, Department of Virology & Liver Unit, Nagoya City University Graduate School of Medical Sciences, Kawasumi, Mizuho, Nagoya 467-8601, Japan.

The development of RNA interference (RNAi)-based therapy faces two major obstacles: selecting small interfering RNA (siRNA) sequences with strong activity, and identifying a carrier that allows efficient delivery to target organs. Additionally, conservative region at nucleotide level must be targeted for RNAi in applying to virus because hepatitis C virus (HCV) could escape from therapeutic pressure with genome mutations. *In vitro* preparation of Dicer-generated siRNAs targeting a conserved, highly ordered HCV 5′ untranslated region are capable of inducing strong RNAi activity. By dissecting the 5′-end of an RNAi-mediated cleavage site in the HCV genome, we identified potent siRNA sequences, which we designate as Dicer-hunting siRNAs (dh-siRNAs). Furthermore, formulation of the dh-siRNAs in an optimized multifunctional envelope-type nano device inhibited ongoing infectious HCV replication in human hepatocytes *in vivo*. Our efforts using both identification of optimal siRNA sequences and delivery to human hepatocytes suggest therapeutic potential of siRNA for a virus.

Hepatitis C virus (HCV) is a major etiological agent that causes chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma. Despite clinical improvements provided by combination therapy with interferon-alpha and ribavirin, this therapeutic approach fails in about half of the patients[1]. Clinical proof-of-concept studies for new therapeutic agents have been reported, and several compounds have progressed into preclinical studies[2,3]. However, drug-resistant viruses appear to emerge readily in response to pharmacological selection by novel protease and RNA polymerase inhibitors. Therefore, the development of new therapies for refractory and diverse HCV genotypes represents a major public health objective.

RNA interference (RNAi) results from sequence-specific post-transcriptional gene silencing by double-stranded RNA[4,5]. The effectors of RNAi are short interfering RNA (siRNA) duplexes (~21–23 nt), which play a key role in the specific degradation of target mRNA. Recent studies have shown that a critical challenge in therapeutic application of RNAi is the identification of potent siRNAs; the functionality of these molecules is affected by the duplex nucleotide base preference and the accessibility of the target RNA[6,7,8,9,10]. Additionally, in the treatment of HCV, the emergence of "escaped/resistant" viruses that harbor point mutations in the target region is a major concern in the potential clinical application of RNAi[11,12,13]. Although current algorithms for the selection of anti-viral siRNAs consist of the guidelines previously derived for conserved sequences of viral genomes[14], there is compelling evidence that one siRNA-resistant human immunodeficiency virus (HIV)-1 had no mutations in the target site of the mRNA, but instead contained a point mutation 7 nt upstream of the target site[15]. This observation means that non-target mutations altering the local RNA secondary structure could ablate the RNAi activity without mutation in the siRNA target sequence. Additionally, there are reports that the secondary structure of target sites in mRNAs strongly reduce siRNA-mediated RNAi activity[16,17], hence the accessibility of certain local target structures on the mRNA is an important determinant in the gene silencing ability of siRNAs[18]. Therefore, the prediction of an effective siRNA targeting to virus genome cannot be identified

in the same fashion. Here, we attempt to predict active siRNA sequences using Dicer recognition in a conserved and highly ordered region of the HCV genome.

Separately, delivering siRNAs to intracellular targets also remains a major obstacle. Over the last several years, significant efforts have been devoted to exploring novel delivery strategies; examples include cationic liposomes and polymer-based nanocarriers[19]. To overcome the problems associated with *in vivo* delivery of siRNAs, both biodistribution to the target organ and intracellular trafficking in target cells of nanocarriers need to be addressed. High siRNA-encapsulation efficacy and uniform particle size also are required. We describe here an improved delivery system consisting of a multifunctional envelope-type nano device (MEND)[20], in which siRNA is encapsulated by cationic charged lipid envelope. To avoid the undesired interaction of cationic MENDs with biological components and subsequent loss of activity, a pH-sensitive property was incorporated into the lipid envelope of MEND by using a novel pH-sensitive cationic lipid, YSK05[21]. For enhanced delivery of cargos into cells, pH-sensitive liposomes have been investigated since the mid-1980s[22,23]. Recently, significant progress has been made in *in vivo* systemic siRNA delivery with lipid nanoparticles (LNPs) composed of ionizable cationic lipids. These LNPs represent neutral surface at physiological pH, but convert to a cationic form under acidic conditions (as expected in the endosome); siRNAs delivered by this mechanism provide efficient reduction of target gene expression in liver[24]. In this study, we describe the development of liver-targeted MENDs containing YSK05 for delivery of the active siRNAs, a system with therapeutic potential for the treatment of HCV-infected liver.

Results

Dicer-hunting siRNA targeting the HCV IRES has powerful silencing efficacy. The most conserved sequences among different HCV genotypes are 5′ untranslated region (UTR). The HCV 5′ UTR forms an RNA folded structure which has functional for the internal ribosome entry site (IRES)[25], and then allows protein synthesis to proceed in a cap-independent manner, implying an important role in key step of the viral translation and replication. Therefore, siRNAs targeting the IRES are expected to reduce the chances of viral mutational escape because the conserved 5′ UTR are likely to contain both structurally and functionally constrained elements (**Fig. 1a**). As the HCV IRES has local higher order structures at the RNA level, random sequence of siRNA targeting to the region might not exclusively induce RNAi activity. To identify an effective siRNA targeting the IRES sequences, we previously tested the efficacy of several synthetic siRNAs using an *in vitro* HCV-replicon assay, revealing that the siE sequences had an IC50 of 167 pM in this assay[26]. In addition, we found that the Dicer-generated siRNAs (d-siRNAs) targeting in the IRES not only provided silencing for heterogeneous target mRNA, but also exhibited even stronger silencing for homogeneous target HCV RNA[26] (**supplementary Fig.1**). Thus, we suspected that d-siRNAs contain powerful siRNA sequences, and/or that d-siRNAs, comprised of a library of several siRNAs, are additive for silencing activity. Previous work has shown that dsRNAs that are longer than 21-mer siRNAs (e.g., 27-mer dsRNAs[27] or 29-mer shRNAs[28]) display enhanced potency in RNAi. We speculated that these longer dsRNAs serve as substrates for the Dicer endonuclease, directly linking the production of siRNA to incorporation into the RNA-induced silencing complex (RISC), for example via the RISC-loading complex[29]. Therefore, we sought to identify the active siRNA sequences in a library of d-siRNAs. Specifically, we screened an siRNA cleavage site of the target HCV genome by using two distinct 5′-rapid amplification of cDNA ends (RACE) methods. The first RACE method employed RNA oligo ligation, such that the 5′-end of a HCV RNA (following cleavage by RISC) was ligated to an RNA oligomer (44 bases); the resulting molecule then was subjected to cDNA synthesis, nested-PCR amplification, and sequencing (**Fig. 1b**). The second RACE method employed C-tailing, such that a series of C nucleotides were attached at the 3′-end of the synthetic cDNA; the resulting molecule was then annealed with an abridged anchor primer and subjected to nested-PCR amplification and sequencing (**Fig. 1c**). We first validated the ability of these RACE methods to detect siRNA-mediated cleavage. Synthetic siE was transfected into REF cells[30] (which harbor the divided-full genome replicon), and total RNA was purified from the transfected cells. Using the two independent 5′-RACE methods, we demonstrated that a unique site, at the 5′-end of the HCV genome, was cleaved following treatment with the siE (**Fig. 1d**). The cleavage site on the target HCV RNA genome was located 10 nt downstream of the 5′-end of the guide siE in 41 of 48 clones. Previous reports show that RISC cleaves the mRNA at a site precisely 10 or 11 nt downstream of the 5′ end of the siRNA guide strand[31]. Thus, our results confirm the utility of RACE for identifying the 5′-end of the cleaved target mRNA during silencing. Next, we determined the 5′-end of HCV genome cleaved by the d-siRNAs, which consist of several kinds of siRNA products generated *in vitro* by activity of the Dicer endonuclease. Treatment with the d-siD5-50 yielded six separate cleavage sites across the 50-nt-long target region of the HCV RNA genome (**Fig. 1e**). Treatment with the d-siD5-197 yielded five separate cleavage sites across the 197-nt-long target region of the HCV RNA genome (**Fig. 1f**). Notably, the si197-1 site, which was observed among the majority of d-siD5-197 clones, was identical to the si50-15 site. Additionally, note that we could not detect the HCV-specific siE site in cloning of d-siD5-50 and d-siD5-197 cleavage sites, although the siE site had exhibited superior silencing efficacy in previous studies[26].

siRNAs, which consist of duplexes of 21-nt RNAs that are base-paired with 2-nt 3′ overhangs, were designed based on the cleavage site defined by d-siRNAs; these siRNAs therefore were designated as Dicer-hunting siRNAs (dh-siRNA). We synthesized a series of 10 individual dh-siRNAs and transfected each into R6FLR-N replicon cells having with the reporter gene (**Fig. 1g** and **1h**). Compared to the efficacy of the siE, several of these individual dh-siRNAs showed strong silencing activity against HCV replication. The efficacy of the si197-1 was three times higher than that of siE; the other two dh-siRNAs silenced HCV replication with efficacy two times higher than that of siE.

Currently, software for efficient siRNA design of antiviral RNAi are available (siVirus[14], siDirect[32], Block-iT (Invitrogen), and the web-based antiviral siRNA design software). These programs make prediction based on highly conserved regions of divergent viral sequences, and claim to minimize off-target effects that might result from effects on the molecular mechanism of RISC assembly or from the sequence preferences of the RISC endonuclease. In the HCV RNA genome sequences, we predicted an effective siRNA targeting the IRES (nt 199-395) using the three commercial software programs. We compared a software-proposed siRNA as previously reported with dh-siRNAs we report here. The values of IC50 for the predicted siRNAs in HCV-replicon assays are presented in **Fig. 1i**. In the IRES region, the siBlock-iT software proposed four kinds of siRNAs; siDirect proposed seven kinds of siRNAs; and siVirus proposed no siRNA. Five kinds of dh-siRNA were proposed by either siBlock-iT or siDirect software, and the dh-si50-11 was proposed by both siBlock-iT and siDirect software. However, the first- and second-most effective siRNAs, dh-si197-1 and dh-si197-6, were proposed by neither siBlock-iT nor siDirect software. These corresponded to molecules with IC50s of 61 and 66 picomolar (respectively) in our replicon assays (**Fig. 1i**). These results indicated that dh-siRNA prepared based on the cleavage site of the target mRNA treated with Dicer-generated siRNAs were powerful siRNA target sites; commercial design software programs were not of service for siRNA prediction in the HCV IRES region. In the case of siRNAs targeting endogenous mRNA, precise processing by Dicer is not

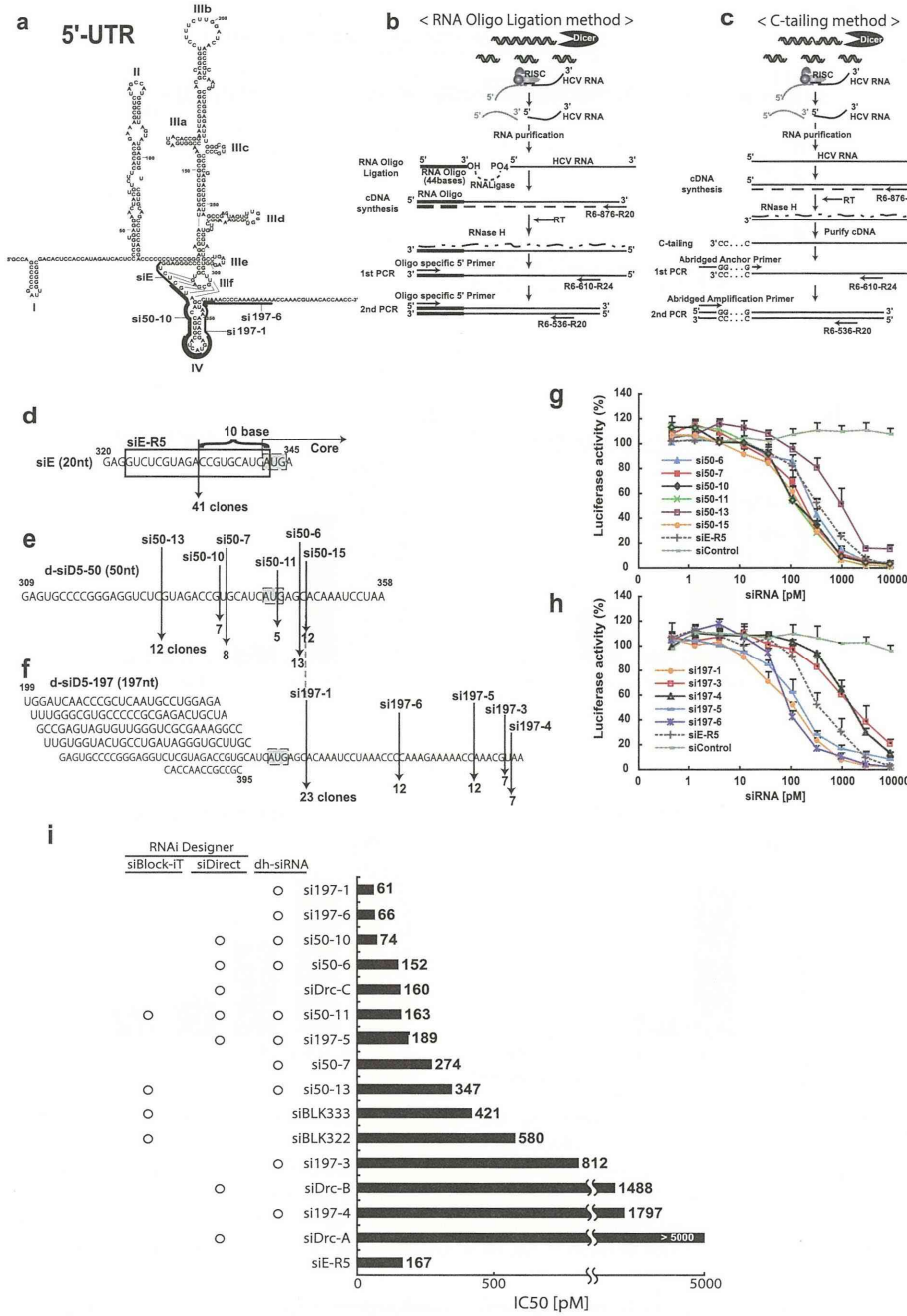


Figure 1 | Generation and efficacy of dicer-hunting siRNAs (dh-siRNAs). (a) Complicated secondary structure of the 5′-UTR in the HCV genome[50] and representation of the potent siRNA sequences we identified (si50-10, si197-1, and si197-6). (b, c) 5′-RACE strategy for identifying the RNAi cleavage sites of the target HCV genome using the RNA oligo ligation method (b) and the C-tailing at the 3′-end of RNA method (c). (d–f) Representation of the cleavage sites by siRNA and the number of clones in the sequences of HCV RNA genome that were used as templates. Cleavage site of the HCV RNA genome by siE (d), d-siD5-50 (e), and d-siD5-197 (f) was identified using the two RACE methods (2b and 2C). (g, h) Evaluation of silencing efficacy of dh-siRNAs. The HCV-replicon cells with reporter genes were transfected with the dh-si50 series (si50-6, 7, 10, 11, 13, and 15) (g) or the dh-si197 series (si197-1, 3, 4, 5, and 6) derived from d-si197 (h). Luciferase activity was measured after 48 hr. Data are presented as mean ± s.d. ($n = 5$) of values normalized to those obtained with mock-transfected cells. (i) Comparison of IC50 for inhibition of HCV replication by siRNAs that were derived from dh-siRNA or predicted by siRNA web design tools. Based on the HCR6 (genotype 1b) sequence, commercial software (siBlock-iT, siDirect) predicted several siRNA sequences (Supplementary table 1). IC50 values represent the mean for independent determinations ($n = 5$) using HCV replicon cells harboring subgenomic HCR6 sequences.

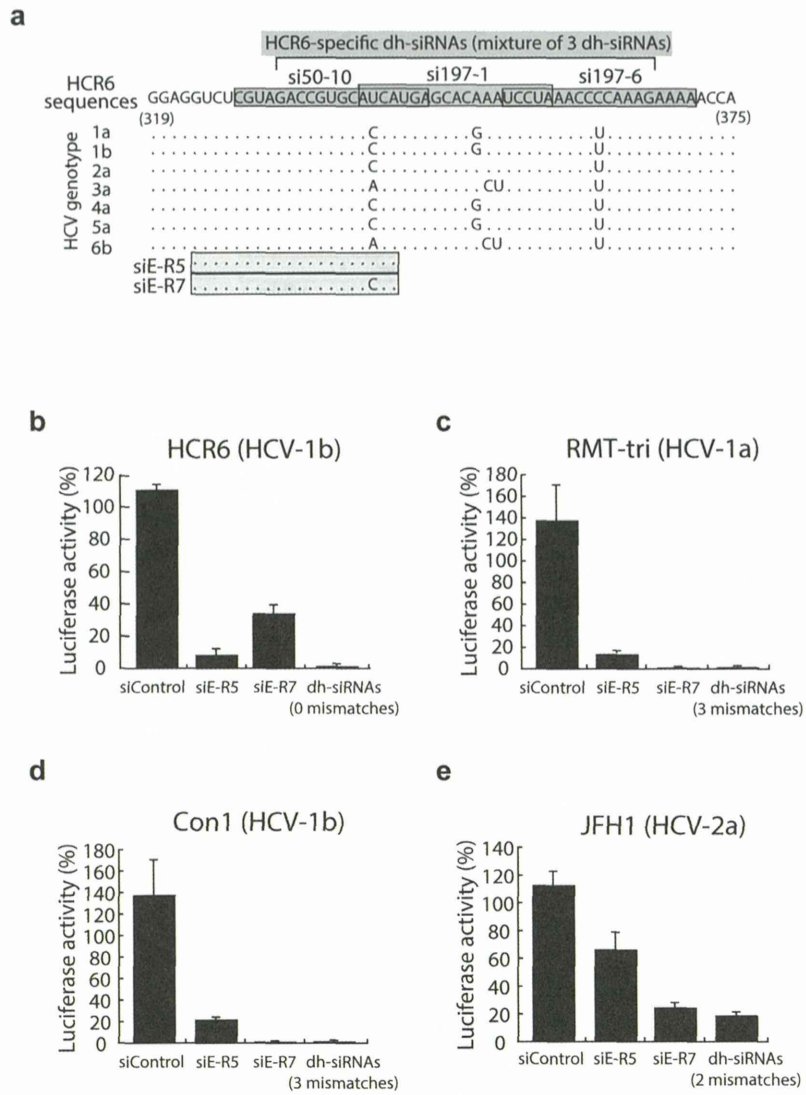


Figure 2 | Efficacy against heterogeneous viruses by mixtures of dh-siRNAs. (a) Alignment of HCV genotype sequences in the region targeted by the dh-siRNAs and siRNA sequences. Aligned sequences for siRNAs are indicated by shaded boxes; dots indicate matches to the consensus (HCR6); individual bases indicate genotype-specific differences. (b–e) Inhibition efficacy by mixture of three HCR6-based dh-siRNAs against homogeneous replicon cells (b), heterogeneous replicon cells with RMT-tri sequences (genotype-1b; 3 mismatches) (c), heterogeneous replicon cells with Con1 sequences (genotype-1b; 3 mismatches) (d), and heterogeneous replicon cells with JFH1 sequences (genotype-2a; 2 mismatches) (e). All samples were assessed at 72 hr after transfection. Data are presented as mean ± s.d. (n = 5) of values normalized to those obtained with mock-transfected cells.

critical because any cleavage frame would result in functional siRNAs; however the correlation of siRNA sequence with structurally/functionally constrained elements of HCV RNA is expected be critical for RNAi function.

Combination of the Dicer-hunting siRNA leads to enhanced reduction for heterogeneous HCV RNA. The notoriously error-prone replication of RNA viruses is a severe challenge for the development of siRNA-based anti-viral therapies. Indeed, HCV displays a high rate of mutation and is classified into distinct genotypes (1 to 6) (**Fig. 2a**). Therefore, to examine whether a mixture of dh-siRNAs can silence the replication of heterogeneous HCV RNA, we transfected HCR6 sequence-specific dh-siRNAs into heterogeneous replicon cells, specifically cells that harbor RMT-tri (genotype 1a), Con1 (genotype 1b), or JFH1 (genotype 2a) replicons (**Fig. 2b–e**). The combination treatment included si197-1, si197-6, and si50-10. All three of these sequences are HCR6-specific (genotype 1b), and the combination silenced the homogeneous HCV replication more effectively than the HCR6-specific siE-R5 alone (**Fig. 2b**). We next transfected a mixture of the dh-siRNAs into RMT-tri replicon cells, which harbor heterogeneous sequences compared to the HCR6 sequences. Treatment with the combination of three dh-siRNAs targeting the HCR6 sequences yielded a single mutation within the RMT-tri (genotype 1a) target sequences (**Fig. 2a**). The combination treatment (dh-si197-1, dh-si197-6, and dh-si50-10) represented a total of 3 mismatches versus the heterogeneous RMT-tri replicon but still provided silencing of heterogeneous RMT-tri replication. For comparison, siE-R5, which represents a single mismatch versus the RMT-tri target sequence, exhibited reduced silencing against the heterogeneous RMT-tri genome (**Fig. 2c**). Moreover, the replication of Con1 (another virus of genotype 1b) and JFH1 (a virus of genotype 2a) were silenced by the dh-siRNA combination treatment, with efficacy equal to or exceeding that of replicon-specific siE-R7 (**Fig. 2d** and **2e**). These results indicate that

a

	Initial MEND	Optimized MEND
Diameter (nm)	105±19	81±3
Polydispersity index	0.12±0.07	0.13±0.04
Zeta-potential (mV)	0.7±4.3	6.1±1.8
Encapsulation efficiency (%)	97±1	98±1
Apparent pKa	6.6	6.4

Data points are expressed as mean±s.d. (n=3).

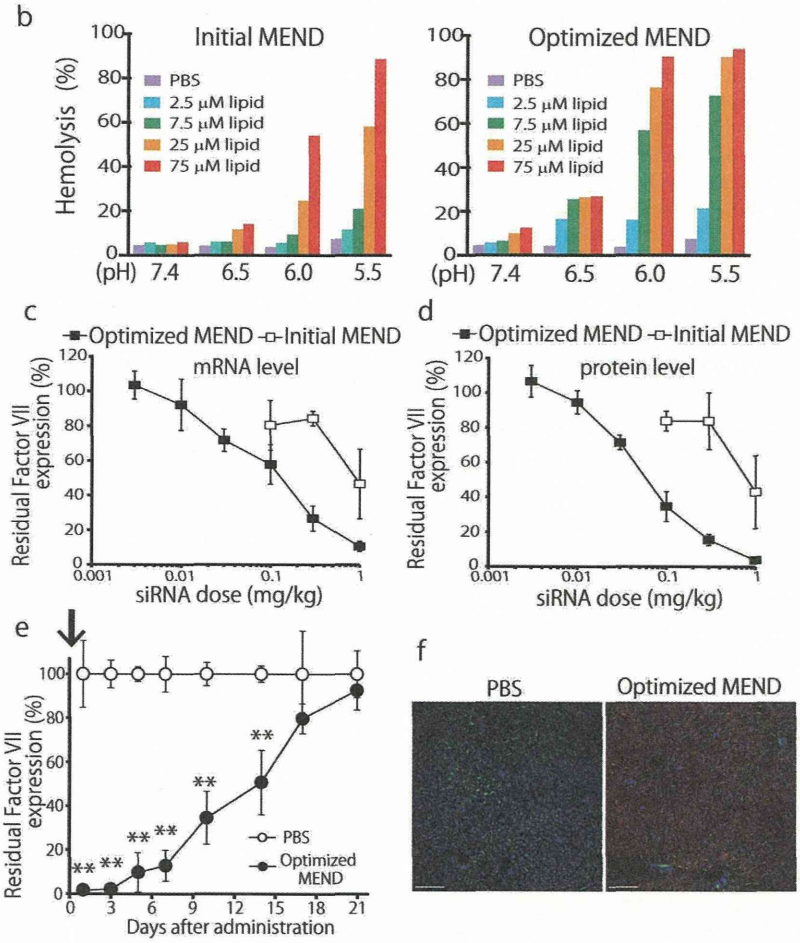


Figure 3 | ***In vitro* and *in vivo* characterization of the optimized MEND.** (a) Physical properties of multifunctional envelope-type nano devices (MENDs). (b) The membrane fusion activity of initial and optimized MENDs was assessed by an *in vitro* hemolysis assay. Values (normalized to hemolysis with known lysing agent (Triton)) are presented as the mean (n=3). N.C.; negative control (treated with PBS). (c, d) *In vivo* hepatic *Factor VII* mRNA (c) and serum Factor VII protein (d) levels at 48 hr after the administration of the initial and optimized MENDs. Data points are presented as the mean ± s.d. (n=3) of values normalized to those obtained with uninjected mice. (e) *In vivo* persistence of knockdown was investigated by monitoring serum Factor VII levels at the indicated number of days after administration of optimized MENDs at a dose of 1.0 mg siRNA per kg. **P<0.01. Data points are presented as the mean ± s.d. (n=3) of values normalized to those obtained with uninjected mice. (f) Liver tissues were collected after single injection of PBS or the optimized MENDs encapsulating Cy5-siRNA (red) and stained with FITC-isolectin B4 (green) and Hoechst33342 (blue) to detect blood vessels and nuclei (respectively). Bars represent 100 μm.

combination treatment with dh-siRNAs targeting the IRES sequence can silence and reduce HCV replication, even when applied to heterogeneous HCV genomes.

Characterization and optimization of pH-sensitive MENDs containing YSK05 for efficient siRNA delivery to hepatocytes. One of the challenges in developing RNAi therapeutics is the efficient delivery of siRNA into the cell affected by a given disease, for instance into the cytoplasm of HCV-infected hepatocytes. As a possible solution, we attempted dh-siRNAs delivery to hepatocytes using a pH-sensitive MEND containing YSK05, a pH-sensitive cationic lipid. The physical properties of siRNAs formulated in MENDs are shown in **Fig. 3a**. To examine the utility of MENDs containing YSK05 for the *in vivo* systemic delivery of siRNA to liver, mice were treated with siRNAs formulated in MENDs. We first screened the lipid composition to achieve efficient knockdown (**supplementary Fig. 2**). We found that MEND composed of YSK05, cholesterol, and PEG-DMG (at ratios of 70 : 30 : 3) provided the most efficient knockdown of target *srbI* gene in liver (**supplementary Fig. 2b**). Next, the membrane fusion activity of the optimized MEND was assessed by hemolysis assay[21] at the indicated pH to evaluate the potency for escape from endosomes via membrane fusion