Identification of Characteristic Genomic Markers in Human Hepatoma Huh7 and Huh7.5.1-8 Cell Lines

Masaki Kawamoto[1], Toshiyuki Yamaji[2], Kyoko Saito[2], Kazuhiro Satomura[3], Toshinori Endo[3], Masayoshi Fukasawa[2], Kentaro Hanada[2,*], and Naoki Osada[3, 4,*]

1. Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
2. Department of Biochemistry & Cell Biology, National Institute of Infectious Diseases, Tokyo 162-8640, Japan
3. Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
4. Global Station for Big Data and Cybersecurity, GI-CoRE, Hokkaido University, Sapporo 060-0814, Japan

*Correspondence

Kentaro Hanada

Email: hanak@nih.go.jp

Address: Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640, Japan

TEL: +81-3-5285-1158

FAX: +81-3-5285-1157

Naoki Osada

Email: nosada@ist.hokudai.ac.jp

Address: Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan

TEL: +81-11-706-7332

FAX: +81-11-706-6546

Abstract

The human hepatoma-derived Huh7 cell line and its derivatives (Huh7.5 and Huh7.5.1) have been widely used as a convenient experimental substitute for primary hepatocytes. In particular, these cell lines represent host cells suitable for propagating the hepatitis C virus (HCV) *in vitro*. The Huh7.5.1-8 cell line, a subline of Huh7.5.1, can propagate HCV more efficiently than its parental cells. To provide genomic information for cells' quality control, we performed whole-genome sequencing of Huh7 and Huh7.5.1-8 and identified their characteristic genomic deletions, some of which are applicable to an in-house test for cell authentication. Among the genes related to HCV infection and replication, 53 genes were found to carry missense or loss-of-function mutations likely specific to the Huh7 and/or Huh7.5.1-8. Eight genes, including *DDX58* (*RIG-I*), *BAX*, *EP300*, and *SPP1* (osteopontin), contained mutations observed only in Huh7.5.1-8 or mutations with higher frequency in Huh7.5.1-8. These mutations might be relevant to phenotypic differences between the two cell lines and may also serve as genetic markers to distinguish Huh7.5.1-8 cells from the ancestral Huh7 cells.

Keywords: Huh7, genome sequencing, cell lines, hepatitis C virus

Introduction

Quality control of cells is crucial for biopharmaceutical manufacturing as well as research. In research, many cell lines have been misidentified (http://iclac.org/wp-content/uploads/ Cross-Contaminations-v8_0.pdf), and various academic journals have begun requesting proper authentication of cells used in studies. Cell lines can be identified by genotyping multiple common genetic variants, such as microsatellite markers and single nucleotide variants (SNVs)(Gilbert et al., 1990, Ghandi, 2019 #11972). However, such common markers would not be useful for discriminating between cell lines that originated from the same ancestral cell line. Thus, rational, easy in-house methods of identifying cell lines are widely desired. In particular, the mutation rate of deletion is much lower than that of point mutations(Wang et al., 2008). Cell line-specific deletions are applicable to cell line authentication of African green monkey kidney-derived Vero cells (Osada et al., 2014)·(Sakuma et al., 2018), which are widely used in virologic research and human vaccine production.

HuH-7 (hereafter Huh7) is a permanent cell line established from male hepatoma tissue, which was surgically removed from a 57-year-old Japanese male in 1982(Nakabayashi et al., 1982). Huh7 and its derivatives have been used as a convenient experimental substitute for primary hepatocytes. Approximately 80% of hepatocellular carcinoma incidents in humans are caused by hepatitis viruses; in all hepatocellular carcinoma incidents, ~25% were caused by hepatitis C virus (HCV) and ~53% were caused by hepatitis B virus (HBV)(Perz et al., 2006). HCV is a positive-stranded RNA virus that infects humans, causing acute and chronic liver diseases. The chronic infection eventually leads to severe symptoms, such as hepatitis and hepatocellular carcinoma. Thus,

67 permanent cell lines suitable for investigating human hepatitis viruses have always been
68 invaluable. After the establishment of Huh7, several studies have attempted to derive cell
69 lines that are more permissive to HCV than Huh7 (Blight et al., 2002; Feigelstock et al.,
70 2010). The whole-genome sequence of Huh7 was determined in a large-scale genome
71 sequencing project of various cancer cell lines (Ghandi et al., 2019); however, a detailed
72 study of the Huh7 genome sequence has not been performed. Huh7.5, a subline of Huh7,
73 was established as a highly permissive cell line to replicate subgenomic and genomic
74 HCV RNA in 2002 (Blight et al., 2002). Interestingly, Huh7.5, but not the ancestral Huh7,
75 has a missense mutation in *DDX58* (or *RIG-I*) gene, which participates in intracellular
76 antiviral defense (Sumpter et al., 2005). Although cell culture systems that recapitulate
77 the HCV life cycle have not been developed for a long time, the JFH-1 HCV strain has
78 been found to produce infective progenitor virions after introduction to Huh7.5
79 cells(Wakita et al., 2005). Huh7.5.1, a subline of Huh7.5 cells, was subsequently
80 established as a cell line with the intent to generate a cell line that cultures JFH-1 more
81 efficiently than Huh7.5 (Zhong et al., 2005). Although Huh7.5.1 displayed faster viral
82 replication kinetics than Huh7, Huh7 and Huh7.5.1 eventually produced a similar level
83 of viral titers (Zhong et al., 2005). We previously found that the expression level of CD81,
84 which is a plasma membrane protein essential for HCV infection (Petracca et al., 2000),
85 was not uniform in the Huh7.5.1 cells, and we thus obtained Huh7.5.1-8, a subclone of
86 Huh7.5.1, in which CD81 is stably expressed on cell surfaces (Shirasago et al., 2015).
87 Huh7.5.1-8 cells exhibit ~10-fold greater permissiveness to HCV than Huh7.5.1 cells
88 (Shirasago et al., 2015). HCV culture systems have provided robust assay systems to
89 screen anti-HCV compounds, resulting in the development of various anti-HCV

90 medicines that are currently marketed, although anti-HCV vaccines have not been
91 developed.

92 We determined the whole-genome sequences of Huh7 and Huh7.5.1-8 cell lines and
93 found characteristic genomic deletions in Huh7 and Huh7.5.1-8, some of which are
94 applicable to an in-house test for cell authentication. In addition, among the genes related
95 to HCV infection and replication, 53 genes were found to carry missense or loss-of-
96 function (LoF) mutations that were not registered in the public germline mutation
97 databases, but were found in the Huh7 and/or Huh7.5.1-8. Among them, 8 genes,
98 including *DDX58* (*RIG-I*), *BAX*, *EP300*, and *SPP1* (osteopontin), contained mutations
99 observed only in Huh7.5.1-8 or mutations with higher frequency in Huh7.5.1-8.

100

101 Materials and Methods

102 *DNA sample preparation and sequencing*

103 DNA samples of Huh7 hepatocarcinoma cell line (JCRB0403) and Huh7.5.1-8 cell line
104 were obtained from the Japanese Collection of Research Bioresources (JCRB) Cell Bank.
105 We confirmed that the HCV replication is slower in JCRB0403 than Huh7.5.1-8 (data not
106 shown). For both samples, fragment libraries (average fragment sizes of ~560 bp) were
107 constructed using TruSeq DNA PCR-Free LT Library Prep Kit (Illumina, San Diego, CA,
108 USA). Paired-end sequences of 150 bp long were determined using HiSeq X (Illumina).
109 Approximately 1.8 billion sequence reads were obtained from each sample. We examined
110 whether viral sequences of HBV and HCV are integrated into the two genomes using

111 VirusFinder 2 software(Wang et al., 2015). The sequences were deposited in the DDBJ
112 DRA database under project ID PRJDB7928.

113 *RNA-seq data*

114 We retrieved previously obtained RNA-seq data from Huh7.5.1 and Huh7.5.1-8
115 (DRR018792 and DRR018793, respectively). Because the initial quality check for the
116 RNA-seq data showed some of the read bases had relatively low quality, we applied
117 Trimmomatic software (version 0.36)(Bolger et al., 2014) to filter out low-quality reads.
118 After trimming low-quality bases of average quality score <20 (window size: 4), reads
119 shorter than 75 bp were filtered out. RNA-seq reads were mapped to the reference human
120 genome using HISAT2 (Kim et al., 2015).

121 *SNV calling*

122 Genomic reads were mapped to the reference human genome sequence (GRCh38
123 primary assembly downloaded from the Ensembl database) using the BWA MEM
124 algorithm (version 0.7.15-r1140)(Li and Durbin, 2009) with a default parameter setting.
125 The read mapping rates of Huh7 and Huh7.5.1-8 were 99.88% and 99.87%, respectively.
126 To generate alignment files for variant calling, GATK Best Practices Pipeline 3.0, which
127 includes duplicated read filtering, realignment around indels, and recalibration of base
128 quality score, was applied to the initial alignments(McKenna et al., 2010).

129 The sites with read depths of ≥14 and ≤100 were used for the following variant calling.
130 The number of sites with ≥14 read depth covered >95% of the reference human genome
131 in our dataset. Initial SNV calling was performed using VarScan (version 2.4.3) (Koboldt
132 et al., 2012) with a base quality score cutoff of 15 and a variant allele frequency cutoff of

133 0.1%. Multiallelic SNVs were removed from the analytical pipeline. Known germline
134 SNVs registered in the public database (dbSNP_149) were filtered out using the
135 SelectVariants program in the GATK package(McKenna et al., 2010). However, when a
136 variant nucleotide in Huh7 and/or Huh7.5.1-8 was different from that in the dbSNP, those
137 variants were kept for further analyses as a novel variant.

138 We defined three categories of newly identified SNVs according to their frequency in
139 Huh7 and Huh7.5.1-8. We tested whether the SNV frequency is higher in Huh7 or
140 Huh7.5.1-8 with statistical significance. Statistical significance was evaluated using the
141 χ^2 test and a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). The SNVs
142 with higher frequency in Huh7 and Huh7.5.1-8 were categorized into "Huh7-predominant"
143 and "Huh7.5.1-8-predominant" categories, respectively. The SNVs that did not pass the
144 criteria were further classified into "shared" SNVs when the variant frequencies in both
145 samples exceeded 0.25.

146 *Structural variant calling*

147 Indels shorter than 50 bp were identified, using VarScan, with the same pipeline for
148 SNV calling. Indels that exactly matched to known germline indels (Mills and 1000G
149 gold standard indels (Mills et al., 2006)) were filtered out from further analyses.
150 Accordingly, short indels were classified into three categories: Huh7-predominant,
151 Huh7.5.1-8-predominant, and shared short indels.

152 Long indels (≥50 bp) were identified using Manta (version 1.1.0) (Chen et al., 2016).
153 As with the SNV calling, we only considered the sites with read depths of ≥14 and ≤100.
154 The estimated structural variant frequency in Manta was difficult when we did not obtain

155 a sufficient number of reads spanning breakpoints. Therefore, we classified the structural
156 variants into the above three categories without considering variant frequencies. Large
157 insertions that match to the insertions in Mills and 1000G gold standard dataset and large
158 deletions that overlapped (≥50% length of the estimated deletion size in the cell lines)
159 with the deletions of Mills and 1000G gold standard and/or 1000 Genome Phase 3
160 structural variants (downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/)
161 were filtered out.

162 *Functional annotation*

163 Functional annotation of mutations was performed using the snpEff software
164 (Cingolani et al., 2012) with an annotation data snpEff_v4_3_GRCh38.86. Using the
165 annotation information, SNVs with strong phenotypic effects (missense mutations,
166 nonsense mutations, and mutations at splicing acceptor-donor sites) were extracted.
167 Functional annotation of genes was performed using the DAVID bioinformatics
168 resource(Jiao et al., 2012). We used the following five fields in DAVID output: disease,
169 functional categories, gene ontology, pathways, and protein domains. The protein-protein
170 association networks were retrieved from the STRING database (Szklarczyk et al., 2015).
171 The Python scripts used in this study were deposited in GitHub
172 (https://github.com/J1011Ibb14029/kwmtools).

173 *Sequence conservation*

174 We investigated the evolutionary conservation at mutated amino acid sites using the
175 amino acid sequences of 12 nonhuman vertebrate species: *Pan troglodytes*, *Pan paniscus*,
176 *Gorilla gorilla, Pongo abelii*, *Macaca mulatta*, *Cricetulus griseus*, *Felis catus*, *Desmodus*

177 *rotundus*, *Phascolarctos cinereus*, and *Gallus gallus.* Sequence data were retrieved and
178 aligned using MEGAX (Knyaz et al., 2018), and sequence alignment was performed with
179 the MUSCLE algorithm (Edgar, 2004).

180 *Experimental validation of deletions*

181 The HeLa cervical carcinoma cell line (JCRB9004) was obtained from the JCRB Cell
182 Bank and used as a control. Genome DNA was prepared from the cell lines using the
183 Blood Genomic DNA Extraction Mini Kit (Favorgen, Ping-Tung, Taiwan, ROC).
184 Deletions in the genomes of Huh7 and Huh7.5.1-8 cells were verified by PCR
185 experiments using a previously reported procedure (Sakuma et al., 2017). Briefly,
186 PrimeSTAR GXL DNA Polymerase (Takara Bio, Otsu, Japan) was used for amplification.
187 The reaction mixture, containing 60 ng genome DNA, was denatured at 98°C for 1 min
188 and then subjected to 40 cycles, consisting of 98°C for 10 s, 61°C for 15 s, and 68°C, for
189 1 min. The amplified products were electrophoresed on an agarose gel and visualized
190 with the Gel Doc EZ imager (Bio-Rad, Hercules, CA, USA). Then, 1 kb Plus DNA
191 Ladder was used as a molecular marker (Thermo Fisher Scientific, Waltham, MA, USA).

192 *Sanger sequencing of DDX58 gene*

193 Genomic DNA was prepared from 1×10^6 cells using the Blood Genomic DNA
194 Extraction Mini Kit (Favorgen) and used as a template to amplify a 446 bp fragment
195 corresponding to bases 30,196 to 30,641 of the human DDX58 genic region (National
196 Center for Biotechnology Information accession number NG_046918.1; 78,023 bp) using
197 PCR. PCR was performed with 35 cycles of amplification (95°C for 10 s, 57°C for 5 s,
198 and 72°C for 5 s) on the PCR Thermal Cycler Dice (Takara Bio) using PrimeSTAR MAX
199 DNA polymerase (Takara Bio), forward primer 5′-

200 GTGGCTTGGTGAAGAATGGGCACAG-3′ (bases 30,196–30,220), and reverse primer
201 5′-CTCAGACTAAGAGGCATGAACTATAAGTGG-3′ (complementary to bases
202 30,612–30,641). The resulting fragments were separated in an agarose gel, purified using
203 with a gel/PCR extraction kit (FastGene; Nippon Genetics, Tokyo, Japan), and directly
204 sequenced by the Sanger method using forward primer 5′-
205 CCCTATTTGGGAAGGTCTGGTGATC-3′ (bases 30,291–30,315) and reverse primer
206 5′-CACTTTTACAGTATTGTCAAGCAGC-3′ (complementary to bases 30,564–30,588)
207 by Eurofins Genomics K.K. (Tokyo, Japan).

208 **Results and Discussion**

209 *Identification of mutations in Huh7 lineages*

210 We obtained 41.0- and 41.1-fold coverages of whole-genome sequence reads from
211 Huh7 (JCRB0403) and Huh7.5.1-8 cells, respectively. No sequences related to HBV and
212 HCV were detected in the genomes using the VirusFinder2 pipeline, as reported
213 previously for HBV using *in situ* hybridization (Tay et al., 1990). To identify novel
214 mutations, we filtered out mutations matched to the previously known germline SNVs
215 and indels from further analyses (see Materials and Methods). These mutations were not
216 found in currently available human mutation database, suggesting that most (even if not
217 all) of these newly-identified SNVs specifically occurred during the establishment of the
218 Huh7 cell line and/or the development of the liver cancer in the patient, from which the
219 cell line was established. For the following analysis, we further classified mutations into
220 three categories: 1) Huh7-predominant mutations, of which frequency is higher in Huh7
221 than in Huh7.5.1-8, 2) Huh7.5.1-8-predominant mutations, of which mutation frequency
222 is higher in Huh7.5.1-8 than in Huh7, and 3) shared mutations, of which frequency is

223 almost equal between the two cell lines. The SNVs of higher frequency in Huh7 than in
224 Huh7.5.1-8, with statistical significance (false discovery rate of 0.05), were classified as
225 Huh7-predominant SNVs and vice versa. The SNVs that did not pass the criteria were
226 further classified into shared SNVs when the variant frequencies in both samples
227 exceeded 0.25. Table 1 summarizes the number of mutations, including SNVs, insertions,
228 and deletions.

229 As expected, many new mutations had higher frequencies in Huh7.5.1-8 than in Huh7
230 (Huh7.5.1-8-predominant class), because Huh7.5.1-8 should have experienced a larger
231 number of cell passages. However, Huh7 also had a substantial number of mutations that
232 were not found in Huh7.5.1-8. The observation was, presumably, accounted for by the
233 fact that the seed stock of Huh7 cells we sequenced was not a direct ancestral cell seed
234 that was used to establish Huh7.5. In total, we identified 394,568 new mutations,
235 including 227,447 SNVs and 29,050 insertions, and 136,930 deletions.

236 Most of the mutations were shared between the two cell lines. These shared mutations
237 were most likely to have arisen before the establishment of the Huh7 cell line.
238 Heterogeneity of mutations is often seen not only among different cancer types but also
239 among cancer cells of different patients diagnosed with the same cancer type. Indeed, the
240 amount and pattern of new mutations in liver cancer tissues were shown to be
241 considerably diverse (Fujimoto et al., 2015). In our study, Huh7 and Huh7.5.1-8 contained
242 a relatively large number of new mutations, although it is impossible to obtain germline
243 cells of the patient from whom Huh7 derived to sequence their genome, and some of the
244 mutations might have occurred in the germline cells. Notably, the number of insertions
245 and deletions was considerably high, and the ratio of indels to SNVs was much higher

246 than the ratio observed in germline cells (Wang et al., 2008). Our validation, using
247 polymerase chain reaction (PCR), confirmed that all six amplified regions contained
248 specific deletions in the cell line DNA, indicating that the false-positive rate would not
249 be substantially high in our variant calling pipeline.

250 Interestingly, we identified one nonsynonymous mutation (K70R) in the *POLD3* gene,
251 which plays an important role in high-fidelity DNA replication (Johansson and MacNeill,
252 2010) in both Huh7 and Huh7.5.1-8. In all 12 nonhuman vertebrates we analyzed, POLD3
253 had lysine residues at site 70, indicating that this amino acid site has a very important
254 function in the protein. *POLD3* mutations have been identified repeatedly in cancer cells
255 (Wang et al., 2014), and they contribute to chromosomal instability (Minocherhomji et
256 al., 2015). We suspected that K70R mutations in POLD3 partly explain the large number
257 of SNVs and indels and the frequent chromosome copy number changes in Huh7(Kasai
258 et al., 2018).

259 *Deletion markers for Huh7 cell lines*

260 To establish efficient deletion markers characteristic of Huh7 cells, we designed PCR
261 primer sets that test the presence of deletion in genomic DNA using the dataset of
262 homozygous deletions identified in whole-genome sequencing. Four long deletions,
263 ranging from 7,400 to 262,700 bp in length, and four short deletions, ranging from 645
264 to 1018 bp in length, were targeted for PCR amplification. We designated regions
265 harboring long and short deletions as DL and DS regions, respectively. In our genome
266 sequencing analysis, Five indels were expected to be present only in Huh7.5.1-8, whereas
267 the others were expected to be shared between Huh7 and Huh7.5.1-8. We used a DNA

268 sample from HeLa cells as a control. Primer sequences and detailed information for each
269 region are presented in Supplementary Table 1.

270 We successfully amplified six regions by genomic PCR and confirmed that they all
271 harbored deletions of expected sizes in Huh7 and/or Huh7.5.1-8, but did not show the
272 signature deletion in the HeLa cells (Figure 1). In total, four deletions (DL1–3 and DS2)
273 were present only in Huh7 and Huh7.5.1-8, and two deletions (DS1 and DS3) were
274 observed only in Huh7.5.1-8. The results are summarized in Supplementary Table 2.

275 *Classification of newly identified SNVs*

276 The newly identified SNVs were further classified into missense mutations, nonsense
277 mutations, mutations at splicing signal sites, and other mutations. In addition, frameshift
278 and large indels harboring protein-coding sequences (including gene fusions) were
279 classified as LoF mutations. Nonsense mutations and mutations at splicing signal sites
280 were also classified as LoF mutations.

281 Among these mutations, we selected genes that have missense and/or LoF mutations.
282 We confirmed that missense mutations are present in the RNA-seq read data, which means
283 that mutated alleles are actually expressed in the cells. In total, 255 and 431 genes were
284 found to have missense and LoF mutations, respectively. The complete list of genes is
285 shown in Supplementary Data 1.

286 We narrowed down the gene list and chose several genes that might be relevant to the
287 HCV replication process. We reviewed previous research and selected keywords that are
288 related to HCV infection and replication (Scheel and Rice, 2013; Bukh, 2016): autophagy,
289 apoptosis, antiviral defense, hepatitis C, innate immune response, and serine protease. In

290 addition, we surveyed the genes involved in protein-protein association networks with
291 nine core genes that showed a strong influence on HCV infection and replication (*CD36*,
292 *CD81*, *CLDN1*, *EGFR*, *EPHA2*, *LDLR*, *OCLN*, *PPIA*, and *SCARB1*).

293 We identified 53 candidate genes, including 12 autophagy-related genes, 22 apoptosis-
294 related genes, 3 antiviral defense genes, 14 HCV-related genes, 8 innate immune-related
295 genes, 4 serine proteases, and 4 genes involved in the HCV core-gene network. Eleven
296 genes were categorized into two different classes. The list of genes and mutations is
297 summarized in Table 2. Two HLA genes, *HLA-DRB1* and *HLA-DRB5*, related to an
298 acquired immune system, contained several LoF mutations shared between Huh7 and
299 Huh7.5.1-8, albeit those genes were not included in the 53 genes in Table 2 since cultured
300 cells lack the acquired immune system. Interestingly, among the 53 candidate genes, 8
301 genes contained Huh7.5.8-1+ mutations, but none carried Huh7-predominant mutations.
302 Among the 8 genes, 4 genes (*BAX*, *COL6A3*, *DEFB104B*, and *SIRPB1*) had LoF
303 mutations, whereas the other 4 genes (*DDX58*, *EP300*, *SPP1*, and *ZNF654*) had missense
304 mutations (Table 2). Future studies will be required to elucidate whether these mutations
305 are relevant to the phenotypes of the Huh7 cell lineage.

306 One of the genes, *DDX58*, which is also known as *RIG-I*, encodes an RNA helicase
307 and has an important function for innate antiviral response (Yoneyama et al., 2004). The
308 mutation is, therefore, a strong candidate for the high permissiveness of HCV in the
309 Huh7.5 lineage. It is noteworthy that our whole-genome sequence approach also
310 identified this gene. However, we found a difference in the mutation frequency between
311 the results of this and previous studies. This incongruence is discussed in the next
312 subsection.

313 Because we do not have sufficient space to discuss all 53 candidate genes for the higher
314 permissiveness of HCV in detail, we here selected three key genes that have important
315 viral replication functions.

316 BAX, which has an LoF mutation observed only in Huh7.5.1-8, forms a dimer with
317 BCL2 and activates apoptosis (Oltval et al., 1993). In the Huh7 cell lineage, the *TP53*
318 gene, which encodes P53 protein and controls the apoptotic process through BAX
319 (Toshiyuki and Reed, 1995), has a frameshift deletion (Table 2). Therefore, BAX may
320 positively regulate apoptosis in Huh7. In contrast, in Huh7.5.1-8, 22% of *BAX* alleles
321 carried frameshift insertions, which would decrease *BAX*'s overall activity. The apoptotic
322 activity, therefore, might have decreased in Huh7.5.1-8 compared with Huh7, which
323 might result in HCV's higher replication efficiency.

324 Another candidate gene, *EP300*, which encodes adenovirus early region 1A-associated
325 protein p300 (Eckner et al., 1994; Ogryzko et al., 1996), harbored two missense mutations
326 (D985G and V1764I) in Huh7.5.1-8. *EP300* has the function of suppressing the
327 autophagy process(Pietrocola et al., 2014). Previous studies have shown that HCV and
328 other flaviviruses replicate in host cells using autophagosomes (Levine et al., 2011;
329 Fahmy and Labonté, 2017). We found that the valine residue at site 1764 was highly
330 conserved among vertebrates. The mutation may have reduced *P300*'s function for
331 autophagy suppression and contributed to efficient HCV replication.

332 The last gene we present here is *SPP1* (osteopontin), which is involved in the
333 remodeling process of bones (Kahles et al., 2014). Previous studies found that (1) the
334 gene expression of *SPP1* in liver tissue was increased according to the progress of
335 hepatitis C (Asselah et al., 2005), (2) *SPP1* enhanced autophagy in human hepatocellular

336 carcinoma cells (Liu et al., 2016), and (3) *SPP1* interacted with HCV proteins and helped
337 replicate and assemble HCV(Iqbal et al., 2018). Indeed, it was reported that *SPP1*
338 expression is up-regulated in Huh7.5 compared with Huh7 (Choi et al., 2014). Our study
339 found that 46% of *SPP1* alleles in Huh7.5.1-8 carried missense mutations of K241T,
340 whereas the frequency of mutation in Huh7 was 0%. The frequency of mutations in both
341 Huh7.5.1 and Huh7.5.1-8 transcripts was close to 50%, indicating the K241T mutation
342 had already been acquired in Huh7.5.1, and the expression of the transcript with threonine
343 allele would not be up-regulated by HCV infection in an allele-specific manner. The
344 missense mutation, as well as the overall elevation of gene expression, may have
345 increased the replication efficiency of the HCV in Huh7 derivatives.

346 *Missense mutation in RIG-I (T55I) was heterologous in the Huh7.5.1-8 genome*

347 As described in the above subsection, we found a T55I mutation in RIG-I that could
348 contribute to the high permissiveness to HCV in Huh7.5.1-8. However, both genome
349 sequencing and RNA-seq data showed the mutation is heterologous in Huh7.5.1-8,
350 contradicting previous findings that the mutation is homozygous in Huh7.5 (Sumpter et
351 al., 2005). Therefore, we resequenced the genomic DNA from Huh7 and Huh7.5.1-8
352 using a Sanger sequencer and verified that the mutation is absent in Huh7 but
353 heterologous in Huh7.5.1-8 (Figure 2). Our recent study showed that a 55T allele
354 transcript harbors an allele-specific large deletion and, presumably, nonfunctional in
355 Huh7.5.1-8 (Saito et al., unpublished results). The results showed that, although T55I
356 mutation is heterozygous in the Huh7.5.1-8 genome, all full-length RIG-I proteins have
357 the T55I mutation.

358 In summary, the whole-genome sequencing of Huh7 and Huh7.5.1-8 provided various
359 genetic characteristics of these cell lines, and some of them are, presumably, applicable
360 to an in-house test for authentication of the Huh7 cell lineage. In addition, 53 genes were
361 found to carry missense or LoF mutations specifically found in the Huh7 cell lineage.
362 Among them, 8 genes contained mutations observed only in Huh7.5.1-8 or mutations
363 with higher frequency in Huh7.5.1-8. These mutations might be relevant to the phenotypic
364 differences between Huh7 and Huh7.5.1-8 and might also serve as genetic markers to
365 distinguish Huh7.5.1-8 cells from the ancestral Huh7 cells.

366 **Acknowledgments**

367 We thank the JCRB Cell Bank for providing DNA samples of Huh7 and Huh7.5.1-8.
368 This work was supported by the Japan Society for the Promotion of Science, KAKENHI
369 (grant 17H04003 to K.H. and grant 18H05511 to N.O.) and AMED-CREST (grant
370 JP18gm0910005j0004 to K.H.).

371 **Author Contributions**

372 Study design: T.E., M.F., K.H., and N.O., performing experiments: T.Y. and K.Sai: data
373 analysis, M.K., K. Sat, and N.O., writing manuscript: M.K., T.Y., K.Sai., K.H., and N.O.

374 **Conflict of Interest**

375 The authors declare no competing interests.

376

377 **Figure Legends**

378 **Figure 1**

379 Evaluation of deletions identified in Huh7 and Huh7.5.1-8 using PCR. Samples in lanes
380 a to c are DNA from Huh7, Huh7.5.1-8, and HeLa cells, respectively. Size markers are
381 indicated by horizontal lines. The designs for PCR amplification and expected results are
382 shown on the bottom right. The region name is given for each gel image. Prefixes DL and
383 DS represent whether the expected deletion length is long (>7.4 kbp) or short (<0.9 kbp).
384 For DL regions, primer pairs used for the experiments are shown below each image. For
385 DS regions, primer pairs s and as1 were used for all regions. Each gel image was cropped
386 from the full-sized gels, but different lanes were not merged into a single image.

387 **Figure 2**

388 Sequence chromatograms of *DDX58* in Huh7 and Huh7.5.1-8. The genomic regions
389 harboring the exon 2 of *DDX58* were amplified from Huh7 and Huh7.5.1-8 DNA samples
390 using PCR and sequenced by the Sanger method. The arrows indicate a base 30,443 on
391 NG_046918.1, and C-to-T transition at this position causes a T55I substitution of the
392 DDX58 amino acid sequence.

393

394 **Tables**

395 **Table 1. Summary of newly identified variants in Huh7 and Huh7.5.1-8**

Mutation type	Huh7-predominant	Huh7.5.1-8-predominant	Shared
SNV	4,094	39,497	183,856
Insertion	911	933	27,206
Deletion	1,898	8,948	126,084

396

397 **Table 2 Mutations potentially contributing HCV permissiveness to Huh7 and**

398 **Huh7.5.1-8**

Gene	Status	Type*	Ensembl transcript (AA type)	Function†
BAX	Huh7.5.1-8-predominant	CID	N/A	2
COL6A3	Huh7.5.1-8-predominant	SSM	N/A	7
DEFB104B	Huh7.5.1-8-predominant	CID	N/A	6
SIRPB1	Huh7.5.1-8-predominant	GFM	N/A	6
ACIN1	Shared	CID	N/A	2
AR	Shared	CID	N/A	5
ATG3	Shared	CID	N/A	1
ATM	Shared	SDA	N/A	1, 2
ATP6AP2	Shared	CID	N/A	2
CD34	Shared	CID	N/A	5
CHMP4A	Shared	CID	N/A	1, 2
CTSA	Shared	CID	N/A	1
EGF	Shared	CID	N/A	5, 4
EIF2AK3	Shared	CID	N/A	5
EIF4G3	Shared	CID	N/A	1
GPX1	Shared	CID	N/A	2

H3CR1	Shared	CID	N/A	5
HTT	Shared	CID	N/A	1, 2
IFNA10	Shared	SSM	N/A	3, 5, 6
IFT20	Shared	SDA	N/A	1
KIR2DL4	Shared	CID, SDA	N/A	5
M6PR	Shared	CID	N/A	4
MAP3K1	Shared	CID	N/A	2
MMP17	Shared	CID	N/A	5
NOTCH4	Shared	CID	N/A	5
OVOS2	Shared	CID	N/A	7
PHLDA1	Shared	CID	N/A	2
PHLPP1	Shared	CID	N/A	2
PRUNE2	Shared	CID	N/A	2
PTEN	Shared	SDA	N/A	2
SNCA	Shared	SDA	N/A	1
TMBIM4	Shared	GFM	N/A	2
TNRC18	Shared	CID, SSM	N/A	7
TP53	Shared	CID	N/A	5, 2, 4
UBE2D3	Shared	CID	N/A	2
DDX58	Huh7.5.1-8-predominant	Missense	ENST00000379883.2 (T55I)	3, 5, 6
EP300	Huh7.5.1-8-predominant	Missense	ENST00000263253.8 (D985G)	1

EP300	Huh7.5.1-8-predominant	Missense	ENST00000263253.8 (V1764I)	1
SPP1	Huh7.5.1-8-predominant	Missense	ENST00000508233.5 (K241T)	5
ZNF654	Huh7.5.1-8-predominant	Missense	ENST00000636215.1 (D1072N)	4
ASAH2	Shared	Missense	ENST00000489640.5 (C74Y)	2
ATP6V0A2	Shared	Missense	ENST00000613625.4 (G82E)	1, 7
C4A	Shared	Missense	ENST00000498271.1 (S347Y)	6
CASP8AP2	Shared	Missense	ENST00000551025.3 (K1245N)	2
CYLD	Shared	Missense	ENST00000427738.7 (T822N)	6
DAB2IP	Shared	Missense	ENST00000408936.7 (R1164C)	6, 2
DIDO1	Shared	Missense	ENST00000266070.8 (S1019C)	2
DPF2	Shared	Missense	ENST00000531989.1 (G41V)	2
IL4R	Shared	Missense	ENST00000563002.5 (A82T)	5
PAK4	Shared	Missense	ENST00000360442.7 (R175W)	2
PIP4K2C	Shared	Missense	ENST00000540759.6 (A300G)	1
TBC1D25	Shared	Missense	ENST00000376771.8 (N277S)	1
TNFRSF10B	Shared	Missense	ENST00000276431.8 (V191A)	5, 2
TP53	Shared	Missense	ENST00000509690.5 (Y88C)	5, 2, 4
TRIM56	Shared	Missense	ENST00000306085.10 (R395W)	3, 6

399 *Classes of LoF mutations: SDA, splice donor/acceptor mutation; GFM, gene fusion
400 mutation; CID, coding sequence InDel; Missense, missense mutation; SSM, start/stop
401 codon mutation.

402 † 1, autophagy-related gene; 2, apoptosis-related gene; 3, antivirus-related gene; 4, genes
403 involved in protein-protein association networks, with nine core genes that showed strong
404 influence on HCV infection and replication; 5, known hepatitis C-related gene; 6, innate
405 immune-related gene; 7, serine protease-related gene.

406

407 **References**

408 Asselah, T., Bièche, I., Laurendeau, I., Paradis, V., Vidaud, D., Degott, C., et al. (2005). Liver
409 Gene Expression Signature of Mild Fibrosis in Patients With Chronic Hepatitis C.
410 *Gastroenterology* 129(6), 2064-2075. doi: https://doi.org/10.1053/j.gastro.2005.09.010.

411 Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and
412 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
413 *(Methodological)* 57(1), 289-300.

414 Blight, K.J., McKeating, J.A., and Rice, C.M. (2002). Highly Permissive Cell Lines for
415 Subgenomic and Genomic Hepatitis C Virus RNA Replication. *Journal of Virology* 76(24), 13001-
416 13014. doi: 10.1128/jvi.76.24.13001-13014.2002.

417 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
418 sequence data. *Bioinformatics* 30(15), 2114-2120. doi: 10.1093/bioinformatics/btu170.

419 Bukh, J. (2016). The history of hepatitis C virus (HCV): Basic research reveals unique features
420 in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control.
421 *Journal of Hepatology* 65(1, Supplement), S2-S21. doi:
422 https://doi.org/10.1016/j.jhep.2016.07.035.

423 Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2016).
424 Manta: rapid detection of structural variants and indels for germline and cancer sequencing
425 applications. *Bioinformatics* 32(8), 1220-1222. doi: 10.1093/bioinformatics/btv710.

426 Choi, Steve S., Claridge, Lee C., Jhaveri, R., Swiderska-Syn, M., Clark, P., Suzuki, A., et al.
427 (2014). Osteopontin is up-regulated in chronic hepatitis C and is associated with cellular
428 permissiveness for hepatitis C virus replication. *Clinical Science* 126(12), 845-855. doi:
429 10.1042/cs20130473.

430 Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program
431 for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6(2),
432 80-92. doi: 10.4161/fly.19695.

433 Eckner, R., Ewen, M.E., Newsome, D., Gerdes, M., DeCaprio, J.A., Lawrence, J.B., et al. (1994).
434 Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein
435 (p300) reveals a protein with properties of a transcriptional adaptor. *Genes & Development* 8(8),
436 869-884. doi: 10.1101/gad.8.8.869.

437 Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high
438 throughput. *Nucleic Acids Research* 32(5), 1792-1797. doi: 10.1093/nar/gkh340.

439 Fahmy, A.M., and Labonté, P. (2017). The autophagy elongation complex (ATG5-12/16L1)
440 positively regulates HCV replication and is required for wild-type membranous web formation.
441 *Scientific Reports* 7, 40351. doi: 10.1038/srep40351

442 https://www.nature.com/articles/srep40351#supplementary-information.

443 Feigelstock, D.A., Mihalik, K.B., Kaplan, G., and Feinstone, S.M. (2010). Increased
444 susceptibility of Huh7 cells to HCV replication does not require mutations in RIG-I. *Virology*
445 *Journal* 7(1), 44. doi: 10.1186/1743-422x-7-44.

446 Fujimoto, A., Furuta, M., Shiraishi, Y., Gotoh, K., Kawakami, Y., Arihiro, K., et al. (2015).
447 Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals
448 hepatitis impact and molecular diversity. *Nature Communications* 6, 6120. doi:
449 10.1038/ncomms7120

450 https://www.nature.com/articles/ncomms7120#supplementary-information.

451 Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., et al.
452 (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 569(7757),
453 503-508. doi: 10.1038/s41586-019-1186-3.

454 Gilbert, D.A., Reid, Y.A., Gail, M.H., Pee, D., White, C., Hay, R.J., et al. (1990). Application
455 of DNA fingerprints for cell-line individualization. *American journal of human genetics* 47(3),
456 499-514.

457 Iqbal, J., Sarkar-Dutta, M., McRae, S., Ramachandran, A., Kumar, B., and Waris, G. (2018).
458 Osteopontin Regulates Hepatitis C Virus (HCV) Replication and Assembly by Interacting with
459 HCV Proteins and Lipid Droplets and by Binding to Receptors αVβ3 and CD44. *Journal of*
460 *Virology* 92(13), e02116-02117. doi: 10.1128/jvi.02116-17.

Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C., et al. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28(13), 1805-1806. doi: 10.1093/bioinformatics/bts251.

Johansson, E., and MacNeill, S.A. (2010). The eukaryotic replicative DNA polymerases take shape. *Trends in Biochemical Sciences* 35(6), 339-347. doi: https://doi.org/10.1016/j.tibs.2010.01.004.

Kahles, F., Findeisen, H.M., and Bruemmer, D. (2014). Osteopontin: A novel regulator at the cross roads of inflammation, obesity and diabetes. *Molecular Metabolism* 3(4), 384-393. doi: https://doi.org/10.1016/j.molmet.2014.03.004.

Kasai, F., Hirayama, N., Ozawa, M., Satoh, M., and Kohara, A. (2018). HuH-7 reference genome profile: complex karyotype composed of massive loss of heterozygosity. *Human Cell*. doi: 10.1007/s13577-018-0212-3.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12, 357. doi: 10.1038/nmeth.3317

https://www.nature.com/articles/nmeth.3317#supplementary-information.

Knyaz, C., Stecher, G., Li, M., Kumar, S., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* 35(6), 1547-1549. doi: 10.1093/molbev/msy096.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., et al. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22(3), 568-576. doi: 10.1101/gr.129684.111.

Levine, B., Mizushima, N., and Virgin, H.W. (2011). Autophagy in immunity and inflammation. *Nature* 469, 323. doi: 10.1038/nature09782.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14), 1754-1760. doi: 10.1093/bioinformatics/btp324.

Liu, G., Fan, X., Tang, M., Chen, R., Wang, H., Jia, R., et al. (2016). Osteopontin induces autophagy to promote chemo-resistance in human hepatocellular carcinoma cells. *Cancer Letters* 383(2), 171-182. doi: https://doi.org/10.1016/j.canlet.2016.09.033.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA

491 sequencing data. *Genome Research* 20(9), 1297-1303. doi: 10.1101/gr.107524.110.

492 Mills, R.E., Luttig, C.T., Larkins, C.E., Beauchamp, A., Tsui, C., Pittard, W.S., et al. (2006). An
493 initial map of insertion and deletion (INDEL) variation in the human genome. *Genome research*
494 16(9), 1182-1190. doi: 10.1101/gr.4565806.

495 Minocherhomji, S., Ying, S., Bjerregaard, V.A., Bursomanno, S., Aleliunaite, A., Wu, W., et al.
496 (2015). Replication stress activates DNA repair synthesis in mitosis. *Nature* 528, 286. doi:
497 10.1038/nature16139

498 https://www.nature.com/articles/nature16139#supplementary-information.

499 Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T., and Sato, J. (1982). Growth of Human
500 Hepatoma Cell Lines with Differentiated Functions in Chemically Defined Medium. *Cancer*
501 *Research* 42(9), 3858-3863.

502 Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The
503 Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. *Cell* 87(5), 953-959.
504 doi: https://doi.org/10.1016/S0092-8674(00)82001-2.

505 Oltval, Z.N., Milliman, C.L., and Korsmeyer, S.J. (1993). Bcl-2 heterodimerizes in vivo with a
506 conserved homolog, Bax, that accelerates programed cell death. *Cell* 74(4), 609-619. doi:
507 https://doi.org/10.1016/0092-8674(93)90509-O.

508 Osada, N., Kohara, A., Yamaji, T., Hirayama, N., Kasai, F., Sekizuka, T., et al. (2014). The
509 Genome Landscape of the African Green Monkey Kidney-Derived Vero Cell Line. *DNA Research*
510 21(6), 673-683. doi: 10.1093/dnares/dsu029.

511 Perz, J.F., Armstrong, G.L., Farrington, L.A., Hutin, Y.J.F., and Bell, B.P. (2006). The
512 contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver
513 cancer worldwide. *Journal of Hepatology* 45(4), 529-538. doi:
514 https://doi.org/10.1016/j.jhep.2006.05.013.

515 Petracca, R., Falugi, F., Galli, G., Norais, N., Rosa, D., Campagnoli, S., et al. (2000). Structure-
516 function analysis of hepatitis C virus envelope-CD81 binding. *Journal of virology* 74(10), 4824-
517 4830.

518 Pietrocola, F., Lachkar, S., Enot, D.P., Niso-Santano, M., Bravo-San Pedro, J.M., Sica, V., et al.
519 (2014). Spermidine induces autophagy by inhibiting the acetyltransferase EP300. *Cell Death And*
520 *Differentiation* 22, 509. doi: 10.1038/cdd.2014.215

521 https://www.nature.com/articles/cdd2014215#supplementary-information.

522 Sakuma, C., Sekizuka, T., Kuroda, M., Kasai, F., Saito, K., Ikeda, M., et al. (2018). Novel
523 endogenous simian retroviral integrations in Vero cells: implications for quality control of a human
524 vaccine cell substrate. *Scientific Reports* 8(1), 644. doi: 10.1038/s41598-017-18934-2.

525 Scheel, T.K.H., and Rice, C.M. (2013). Understanding the hepatitis C virus life cycle paves the
526 way for highly effective therapies. *Nature Medicine* 19, 837. doi: 10.1038/nm.3248.

527 Shirasago, Y., Sekizuka, T., Saito, K., Suzuki, T., Wakita, T., Hanada, K., et al. (2015). Isolation
528 and Characterization of an Huh.7.5.1-Derived Cell Clone Highly Permissive to Hepatitis C Virus.
529 *Japanese Journal of Infectious Diseases* 68(2), 81-88. doi: 10.7883/yoken.JJID.2014.231.

530 Sumpter, R., Loo, Y.-M., Foy, E., Li, K., Yoneyama, M., Fujita, T., et al. (2005). Regulating
531 Intracellular Antiviral Defense and Permissiveness to Hepatitis C Virus RNA Replication through
532 a Cellular RNA Helicase, RIG-I. *Journal of Virology* 79(5), 2689-2699. doi:
533 10.1128/jvi.79.5.2689-2699.2005.

534 Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al.
535 (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic*
536 *Acids Research* 43(D1), D447-D452. doi: 10.1093/nar/gku1003.

537 Tay, N., Chan, S.-H., and Ren, E.-C. (1990). Detection of integrated hepatitis B virus DNA in
538 hepatocellular carcinoma cell lines by nonradioactive in situ hybridization. *Journal of Medical*
539 *Virology* 30(4), 266-271. doi: doi:10.1002/jmv.1890300407.

540 Toshiyuki, M., and Reed, J.C. (1995). Tumor suppressor p53 is a direct transcriptional activator
541 of the human bax gene. *Cell* 80(2), 293-299. doi: https://doi.org/10.1016/0092-8674(95)90412-3.

542 Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., et al. (2005). Production
543 of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Medicine* 11,
544 791. doi: 10.1038/nm1268

545 https://www.nature.com/articles/nm1268#supplementary-information.

546 Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., et al. (2008). The diploid genome
547 sequence of an Asian individual. *Nature* 456(7218), 60-65. doi: 10.1038/nature07484.

548 Wang, K., Yuen, S.T., Xu, J., Lee, S.P., Yan, H.H.N., Shi, S.T., et al. (2014). Whole-genome
549 sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer.
550 *Nature Genetics* 46, 573. doi: 10.1038/ng.2983

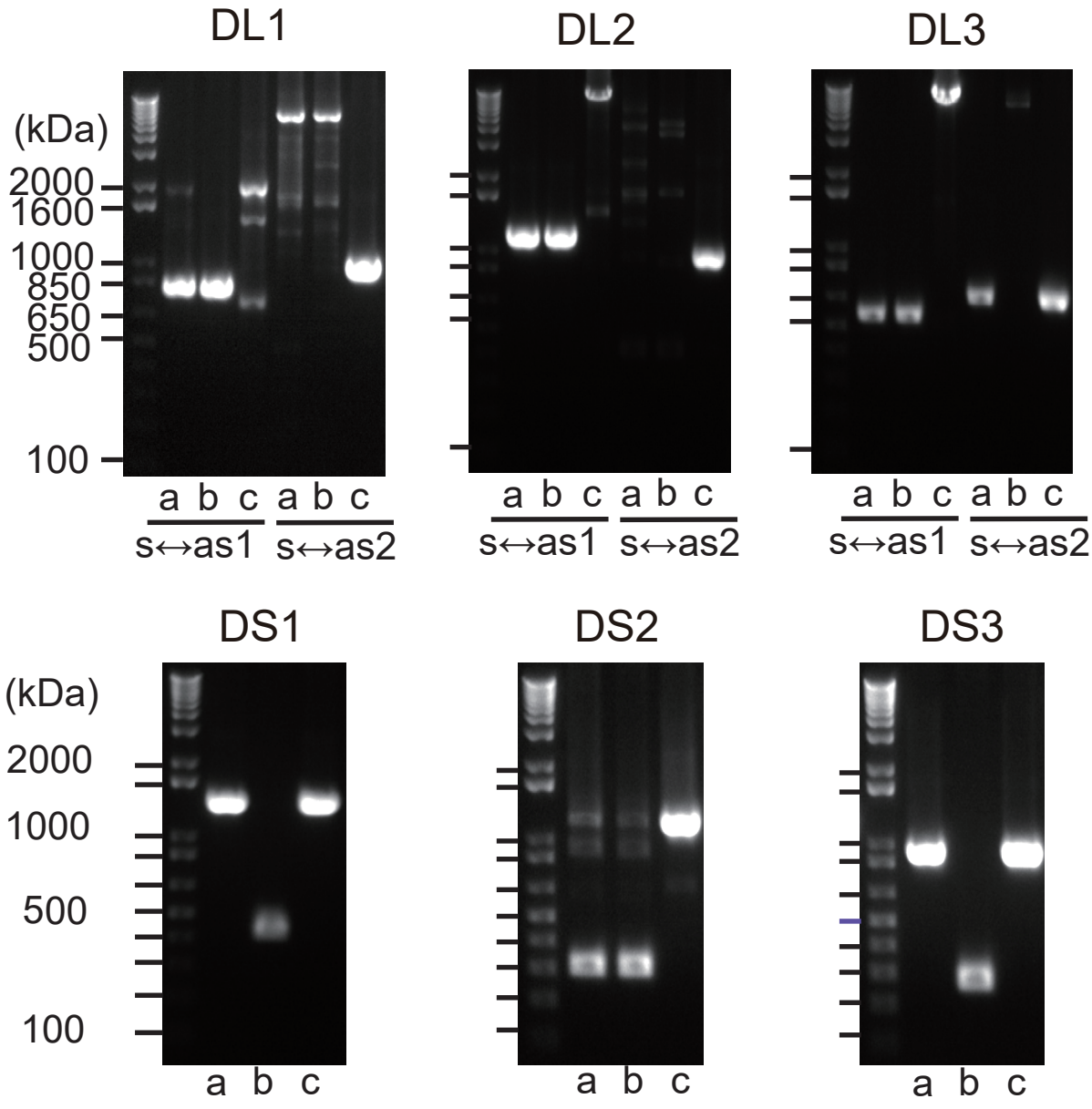
https://www.nature.com/articles/ng.2983#supplementary-information.

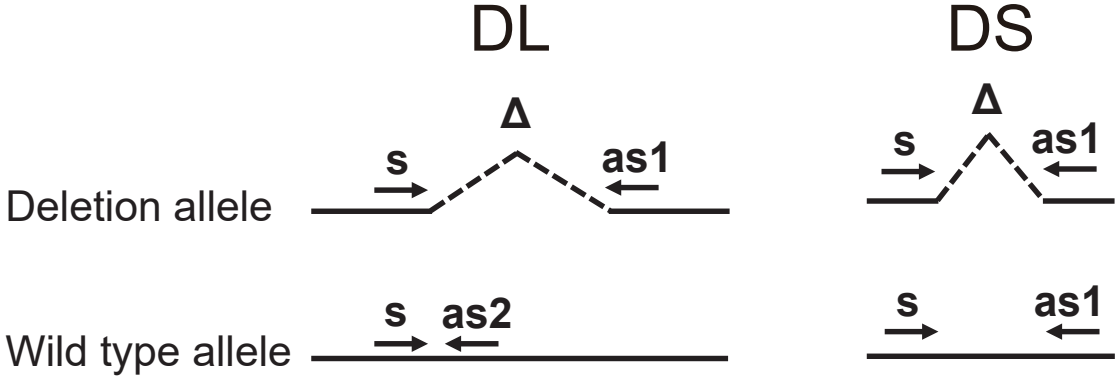
Wang, Q., Jia, P., and Zhao, Z. (2015). VERSE: a novel approach to detect virus integration in host genomes through reference genome customization. *Genome Medicine* 7(1)**,** 2. doi: 10.1186/s13073-015-0126-6.

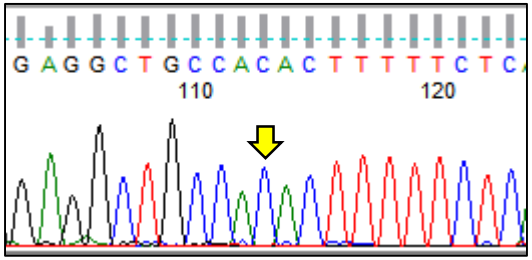
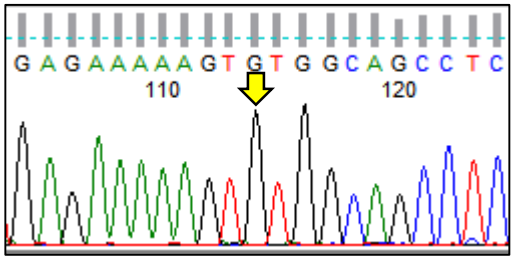
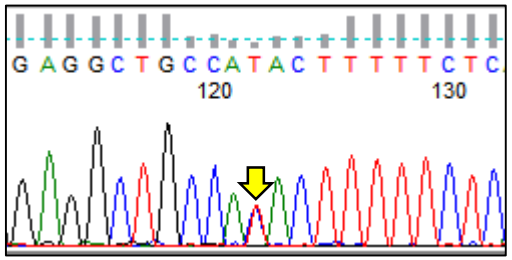
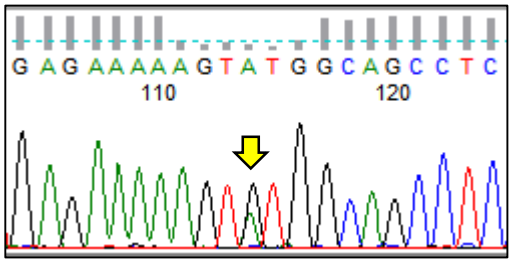
Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., et al. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunology* 5**,** 730. doi: 10.1038/ni1087

https://www.nature.com/articles/ni1087#supplementary-information.

Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D.R., et al. (2005). Robust hepatitis C virus infection in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 102(26)**,** 9294-9299. doi: 10.1073/pnas.0503596102.


DL1
DL2
DL3
(kDa)
2000
1600
1000
850
650
500
100
a b c a b c
s↔as1 s↔as2
DS1
DS2
DS3
(kDa)
2000
1000
500
100
a b c


DL
DS
Δ
s
as1
Deletion allele
s
as2
Wild type allele

	Forward primer	Reverse primer	Base 30,443
Huh7			C
Huh7.5.1-8			C/T