Identification of hepatitis C virus 2k/1b intergenotypic recombinants in Georgia

Mamuka Zakalashvili1 | Jaba Zarkua1 | Michael Weizenegger2 | Jan Bartel2
Monika Raabe2 | Lela Zangurashvili1 | Nino Kankia1 | Nino Jashiashvili1
Maka Lomidze1 | Tengiz Telia1 | Vakhtang Kerashvili1 | Maia Zhamutashvili1
Nikoloz Abramishvili1 | Charlotte Hedskog3 | Krishna Chodavarapu3
Diana M. Brainard3 | John G. McHutchison3 | Hongmei Mo3 | Evguenia Svarovskaia3
Robert G. Gish4 | Irakli Rtskhiladze1 | David Metreveli1

1Hepatology and Gastroenterology Department, Medical Center Mrcheveli, Tbilisi, Georgia
2Medizinisches Versorgungszentrum Dr. Limbach & Kollegen, Heidelberg, Germany
3Gilead Sciences, Inc., Foster City, CA, USA
4Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University Medical Center, Stanford, CA, USA

Correspondence
Mamuka Zakalashvili, Hepatology and Gastroenterology Department, Medical Center Mrcheveli, Tbilisi, Georgia.
Email: mamukazakalashvili@gmail.com

Funding information
This study was funded in part by Medizinisches Versorgungszentrum Dr. Limbach & Kollegen, Heidelberg, Germany, and in part by Gilead Sciences. The preparation of this paper was funded in part by Medical Center Mrcheveli. Writing and editing support was provided by independent medical editor Lark Lands, Ph.D., and funded by Robert G. Gish, M.D.

Handling Editor: Mario Mondelli

Abstract

Background and Aims: This study aimed to evaluate the prevalence of the hepatitis C virus intergenotype recombinant strain RF1_2k/1b in Georgia, confirm viral recombination by full genome sequencing, and determine a genetic relationship with previously described recombinant hepatitis C viruses.

Methods: We retrospectively analysed data from 1421 Georgian patients with chronic hepatitis C. Genotyping was performed with the INNO-LiPA VERSANT HCV Genotype 2.0 Assay.

Results: Virus isolates were assigned to nonspecific hepatitis C genotypes 2a/2c (n = 387) as performed by sequencing of core and NS5B genes. Subsequently, sequencing results classified the core region as genotype 2k and the NS5B region as genotype 1b for 72% (n = 280) of genotype 2 patients, corresponding to 19.7% of hepatitis C patients in Georgia. Eight samples were randomly selected for full genome sequencing which was successful in 7 of 8 samples. Analysis of the generated consensus sequences confirmed that all 7 viruses were 2k/1b recombinants, with the recombination breakpoint located within 73-77 amino acids before the NS2-NS3 junction, similar to the previously described RF1_2k/1b virus. Phylogenetic analysis revealed clustering of the Georgian 2k/1b viruses and RF1_2k/1b, suggesting that they are genetically related.

Conclusions: The 19.7% prevalence of RF1_2k/1b in Georgia patients is far higher than has generally been reported to date worldwide. Identification of recombinants in low income countries with a high prevalence of HCV infection might be reasonable for choosing the most cost-effective treatment regimens.

Keywords: full genome sequencing, hepatitis C virus, phylogenetic analysis, recombinant genotype

Abbreviations: CHC, chronic hepatitis C; HCV, hepatitis C virus; IDU, injecting drug users; NCR, non-coding region; NS, nonstructural; RAV, resistance associated variants; RF, recombinant form; UTR, untranslated region.
INTRODUCTION

It is estimated that 170-180 million people worldwide are chronically infected with hepatitis C virus (HCV) leading to such severe clinical outcomes as cirrhosis and hepatocellular carcinoma and a half million deaths annually.\(^1\)\(^2\) The estimated prevalence of chronic hepatitis C (CHC) in the Republic of Georgia is one of the highest in the world. The last population-based study showed 7.7% seroprevalence in the general population, suggesting 150,000 persons living with HCV infection.\(^3\)

A member of the single-stranded positive-sense RNA virus family Flaviviridae, HCV is characterized by substantial genetic diversity. HCV is currently classified into seven genotypes and many closely related subtypes, with genotypes 1 and 3 dominating in Europe.\(^1\)\(^3\) In 2002, Kalinina et al reported a new circulating HCV strain in St. Petersburg, Russia, which was substantially different from known genotypes.\(^4\) This genetic divergence was due to recombination of HCV genomes of different genotypes rather than to mutations. The authors suggested the addition of the term "recombinant form" (RF) to the current classification of HCV genotypes. Although only the intergenotype recombinant form RF1_2k/1b has been shown to be actively circulating in the general population,\(^5\)\(^-\)\(^8\) other single recombinant isolates of genotypes 2/5, 2b/1b, 2b/1a and 2i/6p have been described and need to be further evaluated clinically and virologically.\(^5\)\(^-\)\(^8\)

During the evaluation of different methods for HCV genotyping at the laboratory Limbach GbR (Heidelberg, Germany) in 2011, four viral variants classified as genotype 2 with the Siemens VERSANT system containing probes targeting the 5′ non-coding (NCR) and core viral genomic region appeared to be 1b with the Abbott Genotyping system targeting the 5′untranslated region (5′UTR) and the nonstructural protein (NS) 5B gene (personal communication, Limbach GbR). With further investigation at the University of Duisburg-Essen in Germany, these HCV isolates were identified as recombinant genotype 2k/1b by sequencing of the core and NS5B regions.

HCV recombination is thought to be rare and to play a minor role in HCV evolution. However, the prevalence of recombination may have been underestimated, given that screening for recombinants is not pursued in routine practice.\(^5\)\(^7\) When "mixed" genotypes are seen, the finding is usually attributed to a "coinfection" with 2 separate variants. In a pilot study in which we evaluated the prevalence of HCV recombinant forms in CHC patients in Georgia, approximately 18.5% were identified as the recombinant form 2k/1b,\(^13\) a prevalence that is far higher than had previously been reported in any other country. To determine the validity of this surprisingly high prevalence, we evaluated the prevalence of this recombinant form in a much larger population of patients and performed phylogenetic analyses to further investigate the genetic relationship with previously described recombinant forms. In addition, because the HCV genotype may affect the outcome of antiviral therapy, variants associated with resistance to direct-acting antivirals were assessed.

MATERIALS AND METHODS

We retrospectively determined HCV genotypes with samples from 1421 CHC patients taken during a 44-month period from October 2012 to May 2016 at the Medical Center Mrcheveli, Tbilisi, Georgia. Approximately 40% of these patients reported a history of injection drug use, and 24% were noted to be active injection drug users. However, in approximately one-third of these patients the route of transmission was unknown. All samples were sent to the Limbach Laboratory (Heidelberg, Germany). We evaluated the prevalence of HCV recombinant forms through core and NS5B sequencing, as well as full genome sequencing for a subset of patient samples. Initially, genotyping was performed using the INNO-LIPA VERSANT HCV Genotype 2.0, a second generation line probe assay containing probes targeting the 5′ non-coding region. For higher resolution

FIGURE 1 Distribution of hepatitis C virus (HCV) recombinant form 2k/1b after sequencing of HCV genotype 2. Sequencing of the core and NS5B region of HCV genotypes 2 (n = 387) by INNO-LIPA showed that 28% (n = 107) of patient samples genotype was classified as 2k, 2a or 2c. In 72% (n = 280) of patient samples, core region sequencing classified the virus as genotype 2k whereas NS5B sequencing classified the virus as genotype 1b, suggesting that the virus was a 2k/1b recombinant form.

Key points

- At 19.7%, the prevalence in Georgia patients of the hepatitis C virus intergenotype recombinant strain RF1_2k/1b is far higher than the ≤3% that has generally been reported to date worldwide.
- The phylogenetic analysis showed that the 2k/1b recombinant viruses from Georgia formed a monophyletic cluster with the previously described RF1_2k/1b sequences.
- The prevalence of this recombinant form may also be found to be elevated in other former Soviet Union countries.
- Identification of recombinants in low income countries with a high prevalence of HCV infection might be reasonable for choosing the most cost-effective treatment regimens.
of genotype 1, the core regions of the viral genome were tested. Refinement of INNO-LiPA genotyping results of all patient samples assigned the unspecified HCV genotypes 2a/2c was performed by sequencing of partial core and NS5B genes. Eight of these patient samples were amplified and sequenced using full genome sequencing, as previously described. With the use of the Ovation RNA-Seq V2 system (NuGEN, San Carlos, CA, USA), RNA was reverse transcribed and amplified. Using the Covaris system (Covaris, Inc., Woburn, MA, USA), the amplified products were fragmented. The Ovation Ultralow DR Multiplex System (NuGEN)
was used to create paired-end libraries for each sample. Deep sequencing of the libraries was done using Illumina MiSeq platform (Illumina, Inc., San Diego, CA, USA). Library preparation, multiplexing and deep sequencing were performed at DDL Diagnostic Laboratory (Rijswijk, Netherlands). Generated sequence reads were processed and assembled by a set of scripts at Gilead Sciences, Inc. Briefly,

**TABLE 1** Presence of nonstructural (NS) 3, NS5A or NS5B resistance associated variants (RAVs) in 2k/1b recombinants from Georgia

| Patient sample | NS3 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|                | V   | F | T | V | Q | S | R | A | D | V | L |   |   |   |   |   |   |   |   |   |
| P257M00004     | 36  | 43| T | 55| 80| 122| 155| 156| 168| 170| 175|   |   |   |   |   |   |   |   |   |
| P257M00005     |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| P257M00006     |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| P257M00008     |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

<table>
<thead>
<tr>
<th>Patient sample</th>
<th>NI</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>S</td>
<td>C</td>
<td>L</td>
<td>V</td>
<td>S</td>
<td>N</td>
<td>C</td>
<td>M</td>
<td>L</td>
<td>R</td>
<td>M</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P257M00004</td>
<td>159</td>
<td>282</td>
<td>289</td>
<td>320</td>
<td>321</td>
<td>96</td>
<td>142</td>
<td>316</td>
<td>any</td>
<td>1/T/V</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>257M00005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P257M00006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P257M00008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The presence of RAVs was investigated in four 2k/1b recombinant sequences, in which NS3, NS5A and NS5B were successfully sequenced. NA, not applicable due to low coverage.
the VICUNA program (Eli and Edythe L. Broad Institute of MIT and Harvard) was used to generate contigs. Using MOSAIK v1.1.0017 the contigs were aligned to subtype-specific references. A full-genome assembly was generated in which reads from each sample were aligned to iteratively improve the full genome consensus sequence. The recombination breakpoint was identified by aligning the generated sequences to the most closely related parental strains of the same subtype. ClustalW was used to align the generated full-genome sequences, and NS3-NS5B regions were extracted and aligned to 93 randomly selected 1b full-length HCV sequences, including two 2b/1b recombinants, two 2b/1b recombinants and 84 non-recombinant genotype 1b sequences obtained from the Los Alamos Hepatitis C Sequence Database (http://hcv.lanl.gov). Maximum likelihood phylogenetic trees were inferred using the open-source GARLI (Genetic Algorithm for Rapid Likelihood Inference, version 2.0) phylogenetic inference program which optimizes the substitution model iteratively. The open-source FigTree (version 1.3.1) graphical user-interface application was used to visualize the phylogenetic trees. As a result of the retrospective nature of this study, ethics approval was not required. All data analysed were collected as part of routine diagnostic procedures and the Health Research Union Institutional Review Board determined that, due to the retrospective nature of this study and the fact that the information obtained would not be linked to identifying information, informed consent would not be required.

## 3 | RESULTS

Distribution of HCV genotypes showed that the most prevalent HCV genotype was genotype 1, 37.5% (n = 532) of all patient samples. Among genotype 1, 94.5% (n = 513) were subtype 1b. Only 5% of patients (n = 17) had subtype 1a; in 2 cases (0.5%), the subtype was not identified. Genotype 3a was the second most prevalent genotype at 34% (n = 483). The unspecific INNO-LiPA genotype 2a/2c was identified in 27% (n = 387) of samples. Genotype 4 was detected in only 0.1% (n = 2). In 1.4% of samples (n = 17), genotyping was unsuccessful.

All patient samples assigned HCV genotype 2 (n = 387) by INNO-LiPA were further investigated by partial gene sequencing of parts of the core and NS5B region. Concordant core/INNO-LiPA and NS5B genotyping results were obtained in 28% (n = 107) of patient samples in which the genotypes were classified as 2k, 2a or 2c. In 72% (n = 280) of patient samples, the core/INNO-LiPA and NS5B genotyping results were discordant; INNO-LiPA/core classified the virus as genotype 2k whereas NS5B sequencing classified the virus as genotype 1b, suggesting that the virus was a 2k/1b recombinant form (Figure 1).

Of the 1421 patient samples, 19.7% (n = 280) appeared to be HCV RF_2k/1b. The majority of the Georgian patients with 2k/1b infection were males and were distributed in all age groups (Figure 2). Our analysis of the distribution of HCV 2k/1b among all HCV genotypes in Georgia showed the high rate of HCV RF_2k/1b; the prevalence of all HCV genotypes identified in this patient population is shown in Figure 3.

Full genome sequencing was performed on eight randomly selected patient samples which were classified as 2k/1b by sequencing of core and NS5B. Seven of eight samples were successfully sequenced and the consensus sequence was generated. Alignment of the consensus sequences together with previously described genotype 2/1 recombinant sequences and non-recombinant genotype 1 and 2 sequences confirmed that these sequences were 2k/1b recombinants. In all seven patients, the recombination breakpoint was

---

### Table 1

<table>
<thead>
<tr>
<th>Q</th>
<th>L</th>
<th>P</th>
<th>S</th>
<th>P</th>
<th>A</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>28</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>38</td>
<td>58</td>
</tr>
<tr>
<td>G/N/R</td>
<td>A/G/T</td>
<td>any</td>
<td>any</td>
<td>L</td>
<td>F</td>
<td>D</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>I</th>
<th>A</th>
<th>V</th>
<th>P</th>
<th>P</th>
<th>A</th>
<th>G</th>
<th>S</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>448</td>
<td>482</td>
<td>494</td>
<td>495</td>
<td>496</td>
<td>499</td>
<td>554</td>
<td>556</td>
<td>559</td>
<td>390</td>
</tr>
<tr>
<td>Any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>any</td>
<td>I</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>RBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
</tr>
<tr>
<td>415</td>
</tr>
<tr>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
</tr>
</tbody>
</table>
4 | DISCUSSION

Our finding that almost 20% of all genotypes and 72% of genotype 2 in Georgia belong to the recombinant form RF1_2k/1b appears to confirm that Georgia has one of the highest prevalences of this recombinant virus so far reported worldwide. Until quite recently, the prevalence of recombinant forms in other countries has not generally exceeded 3% among all HCV genotypes. RF1_2k/1b has previously been reported in patients in Russia (prevalence 2.0%), Cyprus (prevalence 2.6%), Estonia (prevalence 0.5%), Germany (prevalence 1.2%), Uzbekistan (prevalence 1%) and the Netherlands (prevalence 3%).6,8,19-21 Isolated cases have been reported in other countries, including Italy, Azerbaijan, France, Spain and the USA.8,22 Notably, it has been shown that patients infected with this recombinant virus in the former Soviet Union countries were predominantly injecting drug users (IDU), mainly of Georgian origin.21 In the light of this, it was predictable that after the collapse of the Soviet Union further migration could increase the rate of RF1_2k/1b in other parts of the world.22,23 In a very recent study in which sequence analysis was performed with 442 supposed HCV genotype 2 isolates, it was shown that 61 were actually genotype 2k/1b (n = 59); one was found to be 2a/1b and one 2b/1a.23 None of these chimeras were found in samples from patients in Italy, but the frequency was 14% in samples from Germany and 25% in samples from Israel, two countries that have experienced substantial immigration from the former Soviet Union; 88% of the patients with 2k/1b were found to have emigrated from 8 different regions of the Soviet Union.

The RF1_2k/1b strain has been confirmed to originate from a single recombination event dated between 1923 and 1956.8 The fact that the phylogenetic analysis found that the 2k/1b recombinant viruses from Georgia formed a monophyletic cluster with the previously described RF1_2k/1b sequences suggests that they are genetically related to the RF1_2k/1b recombinant virus first described in St. Petersburg in 20024 and are not individual recombination events. The results of our data raise the question of the importance of the prevalence of HCV RF1_2k/1b in the whole Caucasus region. However, because there is currently no data from the neighbouring countries, Armenia, Azerbaijan and Turkey, we do not know if this form is a common HCV genotype outside our tested Georgian population.

Identification of recombinants in high prevalence areas might be reasonable for improving antiviral response rates, particularly in the context of HCV elimination worldwide.14 All currently registered direct-acting antivirals (DAAs) are active only against nonstructural proteins of HCV. It is already known that this particular part of the RF1_2k/1b is similar to genotype 1b so it is suggested that RF1_2k/1b should be treated as genotype 1b. A study by Hedskog et al suggests that the antiviral treatment outcome in patients with the HCV RF_2/1 genotype with a 12-16-week course of sofosbuvir/ribavirin was closer to the response seen with genotype 1 patients than genotype 2 patients.14 In a very recently published study by Susser et al, a very high rate of virological relapse (93%) occurred in RF1_2k/1b patients treated with a genotype 2 regimen, but excellent results were achieved when patients were either initially treated with a genotype 1 regimen (8 of 9 patients achieved SVR) or were retreated after relapse (13 of 13 patients achieved SVR).23

The HCV genotype has long been considered a predictor of the outcome of antiviral therapy. Although the recent development of pan-genotypic DAAs may ultimately eliminate any concern related to the effect of genotype on treatment response, the availability of these very expensive drugs is currently limited in much of the world. Although pegylated interferon/ribavirin has been removed from current HCV treatment guidelines, the use of these regimens still remains (with or without DAAs) in poorer regions and depends on several factors, including regional differences in per capita incomes and health insurance systems.24 It is worth mentioning that six developing nations (China, Pakistan, Nigeria, Egypt, India and Russia) account for more than half of HCV infections worldwide.25 Given the low per capita income in these and other countries, the availability of the pan-genotypic medications may be limited for some time to come, making continuing identification of recombinants potentially important for making treatment choices through whatever time frame is required for the pan-genotypic medications to become widely available and affordable.

In addition, it may be important to identify NS5A Q30R RAVs which in our population were detected in four out of seven RF1_2k/1b patients in which the generated full genome sequences were done. We think that the prevalence of NS5A RAVs in RF1_2k/1b samples and determination of their clinical importance should be evaluated.

In the light of our findings, it is very important to underline the following points: (i) HCV RF1_2k/1b is not unique; (ii) Georgia has one of the highest prevalences of this recombinant virus so far reported worldwide and it appears clear that the prevalence will also be found to be elevated in other former Soviet Union countries and in the countries which have experienced substantial immigration from those countries; (iii) Identification of recombinants in low income countries with a high prevalence of HCV infection might be reasonable for choosing the most cost-effective treatment regimens.

ACKNOWLEDGEMENTS

The authors wish to express deep gratitude for the support received from Professor Robert Thimme, Ärztlicher Direktor, Klinik für Innere Medizin.
CONFLICT OF INTEREST

Dr. Robert G. Gish has had Grants/Research Support from AbbVie, Benitec Biopharma, Gilead Sciences, and Merck & Co. Dr. Gish has performed as Consultant and/or Advisor to AbbVie, Akshaya Pharmaceuticals, AstraZeneca, Bristol-Myers Squibb, Genentech, Gilead Sciences, Hoffman-LaRoche, Ltd., Ionis Pharmaceuticals, Janssen, Merck & Co., Nanogen Biopharmaceutical, and Presidio Pharmaceuticals. Dr. Gish has current activity with the scientific or clinical advisory boards of AbbVie, AstraZeneca, Genentech, Gilead Sciences, Janssen, Merck & Co., and Nanogen Biopharmaceutical. Dr. Gish is a member of the Speakers Bureau for AbbVie, Bristol-Myers Squibb, Gilead Sciences, and Merck. Dr. Gish is a minor stock shareholder of Cocrystal Pharma. Charlotte Hedskog, Krishna Chodavarapu, Mariam M. Holder of Cocrystal Pharma. Charlotte Hedskog, Krishna Chodavarapu, Sibylle Haagmans BL. Identification of a naturally occurring recombinant genotype 2/6 hepatitis C virus. J Virol. 2006;80:7569-7577.

REFERENCES


How to cite this article: Zakalashvili M, Zarkua J, Weizenegger M, et al. Identification of hepatitis C virus 2k/1b intergenotypic recombinants in Georgia. Liver Int. 2017;0:1–7. https://doi.org/10.1111/liv.13540