



OPEN

# Characterization of a new case of XMLV (Bxv1) contamination in the human cell line Hep2 (clone 2B)

Vincent Loiseau<sup>1</sup>, Richard Cordaux<sup>2</sup>, Isabelle Giraud<sup>2</sup>, Agnès Beby-Defaux<sup>3,4</sup>, Nicolas Lévêque<sup>5,6</sup> & Clément Gilbert<sup>1</sup>✉

The use of misidentified cell lines contaminated by other cell lines and/or microorganisms has generated much confusion in the scientific literature. Detailed characterization of such contaminations is therefore crucial to avoid misinterpretation and ensure robustness and reproducibility of research. Here we use DNA-seq data produced in our lab to first confirm that the Hep2 (clone 2B) cell line (Sigma-Aldrich catalog number: 85011412-1VL) is indistinguishable from the HeLa cell line by mapping integrations of the human papillomavirus 18 (HPV18) at their expected loci on chromosome 8. We then show that the cell line is also contaminated by a xenotropic murine leukemia virus (XMLV) that is nearly identical to the mouse Bxv1 provirus and we characterize one Bxv1 provirus, located in the second intron of the pseudouridylate synthase 1 (*PUS1*) gene. Using an RNA-seq dataset, we confirm the high expression of the E6 and E7 HPV18 oncogenes, show that the entire Bxv1 genome is moderately expressed, and retrieve a Bxv1 splicing event favouring expression of the *env* gene. Hep2 (clone 2B) is the fourth human cell line so far known to be contaminated by the Bxv1 XMLV. This contamination has to be taken into account when using the cell line in future experiments.

Continuous cell lines are a cornerstone of cellular and molecular biology as well as of biomedical research. A long-known problem faced by researchers using such cell lines is contamination, whereby foreign cells or microorganisms are inadvertently introduced and remain unnoticed over multiple passages<sup>1,2</sup>. Characterizing cell culture contamination is of prime importance as contaminants may generate flawed experimental results and considerable confusion in the scientific literature<sup>3,4</sup>. The catalog of cell lines misidentified or cross-contaminated with another cell line contains 529 entries as of October 2019<sup>5</sup>. Among the 143 different contaminants listed, the most common is HeLa (118 entries), the first continuous human cell line ever established. HeLa derives from a cervical adenocarcinoma biopsy sampled in 1951 from Henrietta Lacks. Sequencing of the HeLa genome resolved a highly rearranged region on chromosome 8 containing multiple integrated partial genomes of the human papillomavirus 18 (HPV18), thought to have induced tumorigenesis<sup>6</sup>. In addition to standard short tandem repeat (STR) typing, known virus-HeLa cell junctions of HPV18 integrations have been used to unveil cases of cancer cell lines contamination by HeLa<sup>7</sup>.

Multiple microorganisms such as bacteria, fungi, viruses or yeasts have also been identified as persistent contaminants of widely-used cell lines<sup>8–10</sup>. Viral contaminations may be particularly problematic because viruses may present a risk for the persons handling cell cultures and unlike other organisms, viruses cannot be easily detected under light microscopy<sup>9</sup>. Screenings of hundreds of human cell lines have so far revealed only a relatively small proportion of them as contaminated with human viruses<sup>11–13</sup>. Most cases involve contamination by the

<sup>1</sup>Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198 Gif-sur-Yvette, France. <sup>2</sup>Université de Poitiers, CNRS UMR 7267 Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, 5 Rue Albert Turpain, TSA 51106, 86073 Poitiers Cedex 9, France. <sup>3</sup>Laboratoire de Virologie et de Mycobactériologie, CHU de Poitiers, Poitiers, France. <sup>4</sup>Unité de Microbiologie Moléculaire et Séquençage, CHU de Poitiers, Poitiers, France. <sup>5</sup>Laboratoire de Virologie et de Mycobactériologie, CHU de Poitiers, Poitiers, France. <sup>6</sup>EA4331-LITEC, Université de Poitiers, Poitiers, France. ✉email: clement.gilbert@egce.cnrs-gif.fr

Epstein Barr herpesvirus (EBV), either because primary cell cultures were established from infected patients or because EBV was intentionally used to generate transformed lymphoblastoid cell lines.

Non-human viruses have also been found in human cell lines, with murine leukemia viruses (MLV) being responsible for most cases of contaminations<sup>11,13–19</sup>. MLV are gammaretroviruses found under both exogenous and endogenous (proviral) forms in wild and laboratory mouse (*Mus musculus*) strains, which are associated with neoplasias of hematopoietic origin<sup>20</sup>. They are classified in several groups according to their host tropism, which is determined by interactions between their envelope-encoded receptor binding domain and host receptors<sup>21</sup>. While ecotropic MLV interact with the mCAT-1 receptor and can only infect mouse, polytropic MLV interact with the XPR1 receptor and can infect both mouse and other mammalian species. The xenotropic MLV (XMLV) also interact with the XPR1 receptor but they can only infect non-mouse species. Mouse strains are immunized against XMLV thanks to the presence of several *Xpr1* variants that restrict interaction with these viruses<sup>22</sup>. Most cases of human cell contaminations by MLV involve XMLV and it has been shown that contamination by XMLV often leads to important changes in cellular behavior, which can severely affect the conclusions of studies using such infected cell lines<sup>15</sup>.

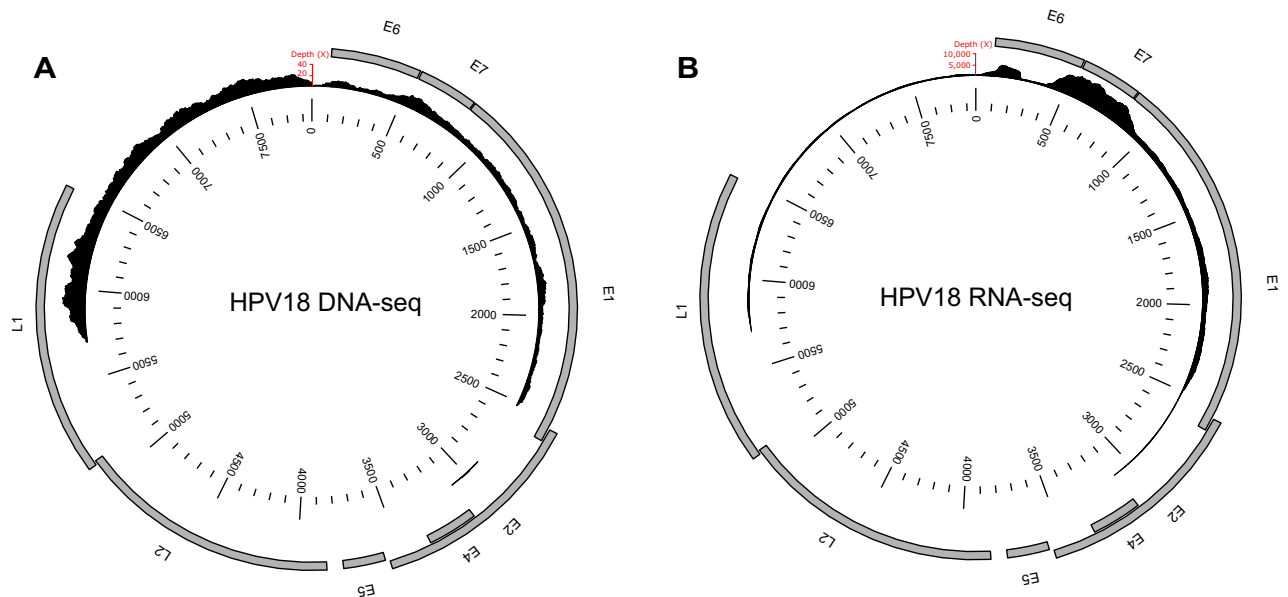
Contamination by XMLV is known to occur either when a human tumor is xenografted in immunodeficient mice to establish or replicate human tumor cell lines, or through unintentional contact between contaminated and non-contaminated cell lines<sup>23–25</sup>. One of the best characterized cases of XMLV contamination is that of the so-called xenotropic murine leukemia virus-related virus (XMRV), which was originally thought to be a possible cause of prostate cancer and chronic fatigue syndrome<sup>26</sup>. XMRV was later shown to result from recombination between two distinct XMLV sequences integrated in the mouse genome (proviruses) and its presence in human cell lines was traced back to a contamination that occurred around 1993, when a prostate cancer cell line (CWR22Rv1) was xenotransplanted into immunocompromised mice<sup>27</sup>. Other well-characterized cases of XMLV contamination involve a virus originating from expression of a mouse chromosome 1 proviral locus called the B10 xenotropic virus 1 (Bxv1)<sup>28</sup>. This provirus is present in many severe combined immunodeficient (SCID) and nude mice strains<sup>20</sup> and Bxv1 infection of human cell lines passed onto these mice strains has been experimentally demonstrated<sup>24</sup>. So far, Bxv1 contamination has been detected in two prostate cancer cell lines, one B-cell line and two pancreatic  $\beta$  cell lines, in which it was shown to produce infectious viral particles<sup>16,17,29</sup>.

Here, we report a new case of a human cell line (Hep2[clone2B], Sigma-Aldrich catalog number: 85011412-1VL) contaminated by Bxv1 using high-throughput RNA- and DNA-seq. As part of a study aiming to characterize the potential presence of infectious agents in the cell lines of our laboratory, we set out to sequence total RNA and DNA extracted from Hep2 (clone 2B) cells. We characterized sequences from two viruses, the HPV18 papillomavirus and the Bxv1 XMLV. The presence of the former is consistent with the known contamination status of Hep2 (clone 2B) cell line by HeLa and the presence of the latter most likely results from mixing with a cell line or reagents contaminated with this virus.

## Materials and methods

**Cell cultures and DNA- and RNA-seq.** Hep2 (clone 2B) cells were purchased from Sigma-Aldrich (catalog number: 85011412-1VL) on June 19, 2013 and cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin (Pen-Strep; Life Technologies) at 37 °C in a 5% (vol/vol) CO<sub>2</sub> atmosphere. Upon confluence, cells contained in a T75 flask were harvested in a 15 ml Falcon tube after trypsinization, washed in Earle's balanced salt solution (EBSS) and centrifuged at 1100g for 5 min. Total RNA and DNA were then extracted using the AllPrep DNA/RNA Mini Kit (Qiagen). A paired-end DNA library (mean insert size 191 bp) was constructed with the DNA sample and a stranded RNA library was constructed with the RNA sample after rRNA depletion. Both libraries were tagged and sequenced on an Illumina HiSeq2500 platform in a 2 × 125 bp configuration in High Output mode (V4 chemistry). Raw DNA- and RNA-seq fastq reads generated for this study have been deposited in dbGaP under accession number phs001944.v1.p1. All reads mapping onto the Bxv1 (JF908815) genome are provided in Supplementary Data 1 and 2.

**Bioinformatics analyses.** Demultiplexing was performed by the sequencing company (Genewiz), yielding 140,313,660 reads with a mean quality score of 35.15 for the DNA sample and 57,593,594 reads with a mean quality score of 35.82 for the RNA sample. Remaining adapter sequences were removed and reads were trimmed with Trimmomatic using default parameters (“-phred33, Illuminaclip:adapter\_file.fasta:2:30:10, minlen:126”, Bolger et al. 2014). To identify infectious agents potentially present in the cells, we mapped paired reads onto the human genome (UCSC genome data, human genome version hg19) using Bowtie2 in “sensitive-local” mode (Langmead and Salzberg 2012). Reads that did not map onto the human genome were assembled for each dataset. DNA-seq and RNA-seq reads were assembled with Masurca<sup>30</sup> and Trinity<sup>31</sup>, respectively, using default settings. The resulting contigs were used as queries to perform blastn and blastx searches on the non-redundant (NR) database of GenBank. To assess sequencing depth of the two viruses identified in this study (Bxv1 XMLV and HPV18), the reads were mapped onto the genome of these viruses (accession number: JF908815 for Bxv1 and GQ180792 for HPV18) with Bowtie2 in “sensitive-local” mode. To identify recombination breakpoints within each of the two genomes and between the two genomes and the Hep2 (clone 2B) genome we used all reads as queries to perform separate blastn searches (megablast option) on the human genome and on the Bxv1 and HPV18 genomes. Virus-virus and cell-virus junctions were then searched within reads using the pipeline described in Ref.<sup>32</sup>. Briefly, only reads aligning over at least sixteen bp on a genome region only and over at least sixteen bp on another genome region only were retained. Reads had to align on at least 100 bp of their length. The overlap between alignment on the virus and on the host sequences was set to involve at most 20 bp and at least -5 bp (see Supplementary Fig. S6 in Gilbert et al.<sup>32</sup>).



**Figure 1.** HPV18 sequencing depth by DNA-seq (A) and RNA-seq (B) reads. The coverage patterns indicate that a portion of the E2 and L1 genes as well as the entire E4, E5 and L2 genes are missing and that the E6 and E7 oncogenes are highly expressed.

**Checking for contamination by rodent DNA.** To assess the presence of rodent DNA or RNA in our samples, reads were mapped using Bowtie2 in “end-to-end” mode on the latest genome version of *Mus musculus* (GRCm38/mm10) and *Rattus norvegicus* (GRSC 6.0/rn6). Using the SAMtools depth program, the mean rodent genome coverage as well as the percentage of the genome covered by the reads were determined. The nature of regions covered by more than 8000 reads was checked in the UCSC genome browser.

**PCR verifications.** To rule out the possibility that the Bxv1 contamination occurred in our laboratory, we purchased a second batch of Hep2 (clone 2B) cells from Sigma-Aldrich in May 2018 and searched by PCR for the presence of Bxv1 in this new batch and in the Hep2 (clone 2B) cells we used for the sequencing (batch ordered at Sigma-Aldrich in 2014). Amplification reactions were performed from 5 ng of DNA extracted from Hep2 (clone 2B) cells by using 10  $\mu\text{mol L}^{-1}$  of each primer (Bxv1\_1-F: AAGAGAAAGAGAGGGACCGC; Bxv1\_1-R: TTTCCTCCAGTAGCCCCTTG), 3 mM MgCl<sub>2</sub>, 0.2 mM dNTP and 0.75 unit of DreamTaq DNA polymerase (Thermo Fischer Scientific) under a 35-cycle PCR program (95 °C for 4 min; 35 cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 15 s, and 72 °C for 10 min). Then a migration of the PCR products was performed on a 1.5% agarose gel at 100 V during 25 min and bands were visualized with a Bio-Rad Transilluminator Universal Hood II under UV light. To check whether regions of the Bxv1 genome not covered by our DNA-seq dataset were in fact present in Hep2 (clone 2B) cells, we PCR-amplified and Sanger-sequenced three such regions as well as another region covered by our DNA-seq dataset using the following primer pairs: Bxv1\_2-F: CCCCAGAAGAGAGAGAAGAAC; Bxv1\_2-R: CATTGGTCCTTATCGAGTTGG; Bxv1\_3-F: TGCCTTTGAGTGGAGAGATC; Bxv1\_3-R: CTAGGGTTTGTAGAAGGGCC; Bxv1\_4-F: CCTTCTCAACAACCTGGGAC; Bxv1\_4-R: ACAGGTCAGCTTGTGTTG; Bxv1\_5-F: CAGGCAAGCTAACTATGGGA; Bxv1\_5-R: CCCAGATTACCTCGGTTTCA.

## Results and discussion

### Confirmation of Hep2 (clone 2B) contamination by HeLa based on HeLa-HPV18 characteristics.

As mentioned on the Sigma Aldrich website, the Hep2 (clone 2B) cell line (catalog number: 85011412-1VL) was originally derived “from tumours produced in irradiated-cortisonised weanling rats after injecting with epidermoid carcinoma tissue from the larynx of a 56-year-old male, but it was later found to be indistinguishable from HeLa by STR PCR DNA profiling.” It is a typical case of non-existent cell line that may cause important problems in cancer research<sup>33</sup>. Consistent with HeLa contamination, assembling of RNA- and DNA-seq reads obtained from sequencing this cell line and not mapping on the hg19 human genome yielded various contigs that were almost identical to the HPV18 (GenBank accession number: NC\_001357.1). As previously reported<sup>7</sup> for HPV18 sequences integrated into the HeLa genome, mapping of both RNA- and DNA-seq reads onto the HPV18 genome shows that a portion of the E2 and L1 genes as well as the entire E4, E5 and L2 genes are missing (Fig. 1). Moreover, the mapped regions displayed the same 23 SNPs between the reads and the HPV18 genome than those identified between HeLa-HPV18 and HPV18 (Cantalupo et al.<sup>7</sup>) (Supplementary Table 1). Mean DNA-seq depth varies from 1.3 $\times$  for the partial E2 gene up to 26.4 $\times$  for the L1 gene (Fig. 1A), reflecting the complex and partially duplicated structure of the integrants<sup>6</sup>. Mean RNA-seq depth varies from 137 $\times$  for the partial E2 gene up to 4078 and 8093 $\times$  for the E6 and E7 genes, respectively. This is in agreement with the known

Number of DNA-seq reads supporting the junction	Number of RNA-seq reads supporting the junction	Viral breakpoint position	Viral gene	Position of breakpoint in human genome	Human chr. Band
1 <sup>a</sup>	272	929	E1	128241377	8q24.21
25	97	5735	L1	128230628	8q24.21
1	57	2497	E1	128241551	8q24.21
1	15	930	E1	128231213	8q24.21

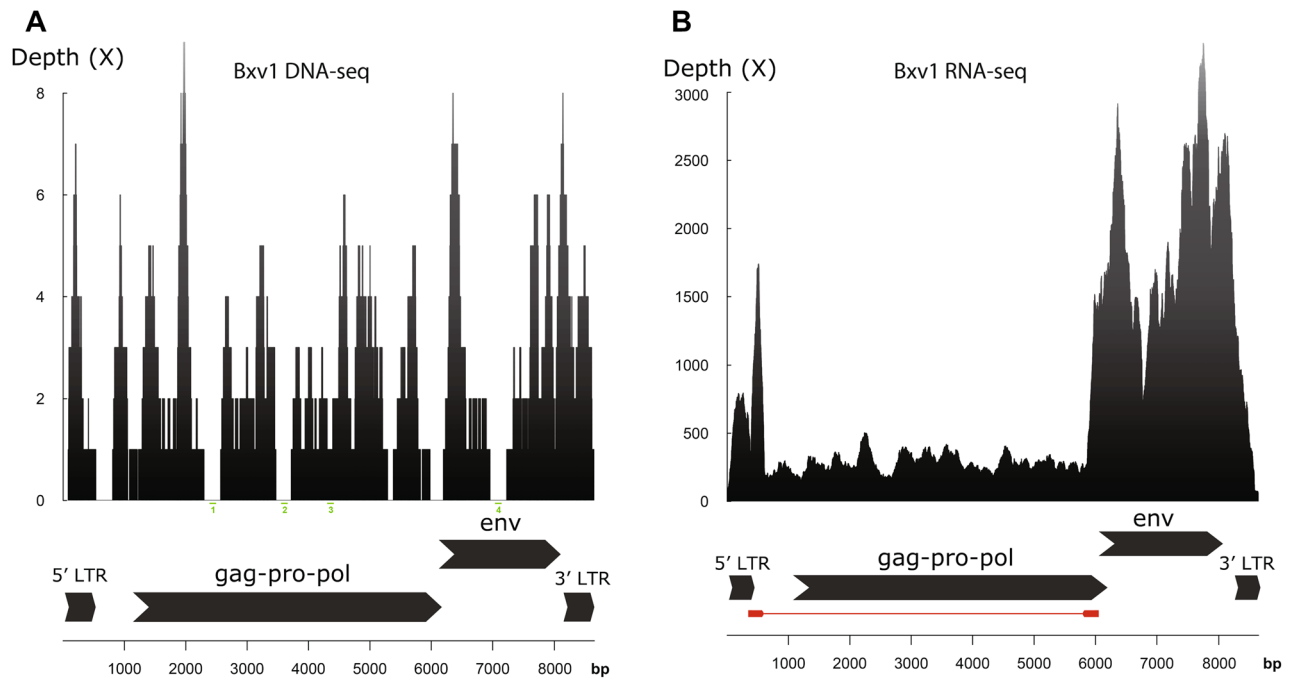
**Table 1.** Characteristics of HeLa–HPV18 junctions supported by both DNA- and RNA-seq reads. <sup>a</sup>Several steps involved in the construction of an Illumina library (including cDNA library synthesis and illumina PCR) may generate artificial chimeras<sup>46,47</sup>. Thus, relying only on one read to identify a breakpoint is not good practice. However, this Table only reports HPV18–HeLa breakpoints that are supported by reads generated in two independent sequencing experiments (RNA-seq and DNA-seq), including one in which they are supported by multiple reads (here more than 15). For example, the first lane of the Table describes a breakpoint between position 929 of the HPV18 genome and position 128 241 377 of human chromosome 8 that is retrieved in 1 DNA-seq read and in 272 RNA-seq reads independently. The position of all breakpoints reported here is consistent with those reported in earlier studies (see text for details).

high expression of the latter two genes, which are involved in oncogenesis through neutralization of tumour suppressors<sup>6,34,35</sup>. Finally, our search for integration loci in DNA-seq reads revealed four virus-cell junctions supported by one or more reads and also supported by RNA-seq reads (Table 1; Supplementary Table 2). All four junctions fall within the chromosome 8 region where HPV18 is known to be integrated in HeLa cells (8q24.21; between positions 128,228,000 and 128,243,000)<sup>6,7</sup>. Much like in Cantalupo et al.<sup>7</sup>, we also identified a number of other junctions in RNA-seq reads, among which all those supported by more than one reads fall within the 8q24.21 region (Table 1; Supplementary Table 2). As noticed by Cantalupo et al.<sup>7</sup>, many of the virus-cell RNA junctions involve the 929 5' splice site in E1 of HPV18, likely indicating that after read-through transcription, splicing events fused the 929 5' donor to a downstream acceptor site located in the human genome<sup>7,36</sup>. Altogether, these results confirm that Hep2 (clone 2B) was contaminated by HeLa at some point during its propagation and they show that HPV18 sequences integrated in this subclone of HeLa have the same characteristics as in other subclones in terms of position in the genome and expression pattern<sup>6,7,34</sup>.

**Hep2 (clone 2B) (HeLa) cells are contaminated by Bxv1 XMLV sequences.** In addition to HPV18 integrants, our assembly of non-human reads also yielded contigs 100% identical to Bxv1, an XMLV proviral locus known to generate infectious particles able to infect human cells and so far identified as a contaminant in four human cell lines<sup>17,20,29</sup>. As several studies have shown that the presence of XMLV in human cell lines could be due to contamination by mouse DNA<sup>19,37</sup>, we checked for such contamination by mapping both DNA-seq and RNA-seq reads onto the mouse genome. Given that the original Hep2 cells were obtained by passaging larynx carcinoma cells onto immune-compromised laboratory rats<sup>38,39</sup>, we also monitored the possible presence of rat DNA by mapping all reads onto the rat genome. Only  $\approx$  4% of the reads mapped onto the rodent genomes, with only  $\approx$  1% of the genome covered and a mean sequencing depth of 0.26 $\times$  for the two species. Importantly, mapped regions corresponded exclusively to RNA genes (7SK, 7SL, U1, U2) that are known to be highly similar between mammalian species and no read was found to map onto intracisternal-A particle elements, which are rodent-specific transposable elements typically used as markers of contamination<sup>19,37</sup>. Thus, the presence of Bxv1 is not due to contamination by rodent DNA but rather results from mixing with a contaminated cell line or reagent. To further verify that contamination did not occur in our laboratory subsequent to reception of Hep2 (clone 2B) cells on June 19 2013 we purchased a second batch from Sigma-Aldrich on May 11, 2018 and validated the presence of Bxv1 sequences by PCR (Supplementary Fig. 1). We conclude that Hep2 (clone 2B) cells from Sigma-Aldrich (catalog number: 85011412-1VL) are not only undistinguishable from HeLa but they are also contaminated by Bxv1 XMLV sequences.

The finding of Bxv1 in the HeLa-contaminated Hep2 (clone 2B) cell line begs the question of whether the virus is present in HeLa cells. XMLVs have been searched in a large number (> 600) of human cell lines, including HeLa and nine of its derivative subclones in which it was found to be absent<sup>18,19,29</sup>. In agreement with this, our blastn search of the Bxv1 sequence onto two HeLa short read datasets available in the SRA database of NCBI (accession numbers: SRX699196 and ERX3188263) revealed no significant hit. This indicates that Bxv1 is likely absent from all or most HeLa cell lines. However, we cannot exclude that this virus was present in the very HeLa cells that contaminated Hep2 (clone 2B). Thus, whether Bxv1 was introduced into Hep2 (clone 2B) by HeLa cells, or whether it was introduced into Hep2 (clone 2B) before or after this cell line was contaminated by HeLa, remains an open question.

**Characterization and expression of a Bxv1 provirus.** To further validate the presence of Bxv1 in the HeLa-contaminated Hep2 (clone 2B) cell line, we searched for evidence of integration of the retrovirus. We identified two virus-cell junctions covered by more than one DNA-seq read and/or also supported by an RNA-seq read (Supplementary Data 3). Interestingly, the two junctions involve the very first and very last position of the long terminal repeat of Bxv1 and the same position in the human genome, suggesting that they correspond to the 5' and 3' extremities of one same proviral locus. Alignment of the junctions with the corresponding region of the human genome revealed that Bxv1 generated a 5-bp target site duplication (AAACC) upon integration (Sup-



**Figure 2.** Bxv1 sequencing depth by DNA-seq (A) and RNA-seq (B) reads. The coverage patterns indicate that the entire Bxv1 genome is present in Hep2 (clone 2B) cells and that the *env* ORF is more expressed than the *gag-pro-pol* ORF. The green lines in (A) show the position of viral regions that were Sanger-sequenced. Numbers below the green lines correspond to those used to name Sanger-sequencing reads in Supplementary Data 4. The red line in (B) represents the spliced region between positions 587 and 5887.

plementary Data 3). Much like Bxv1 proviruses from two prostate cancer cell lines (LAPC4 and VCaP) which all integrated into introns<sup>17</sup>, the Bxv1 provirus characterized here lies within the second intron of the pseudouridylylate synthase 1 (*PUS1*) gene. Furthermore, in agreement with the known propensity of MLV to preferentially target transcription start sites (TSS) and CpG islands<sup>40,41</sup>, the Hep2 (clone 2B) Bxv1 lies only 946 bp downstream of the nearest *PUS1* TSS, well within a 1463-bp long CpG island. Altogether, these results indicate that the Bxv1 integration identified in Hep2 (clone 2B) cells is a *bona fide* proviral locus, confirming contamination by this virus of the Hep2 (clone 2B) cell line.

To assess how many proviral loci may segregate in the cell line, we mapped all reads on the Bxv1 genome (accession number: JF908815). While several short segments amounting to 15% of the Bxv1 genome are not covered by DNA-seq reads (Fig. 2A), we believe this is due to stochastic under-representation rather than true absence of some regions because Bxv1 is fully covered by RNA-seq reads (Fig. 2B). In agreement with this, we were able to PCR-amplify and Sanger-sequence three regions not covered by DNA-seq reads (Fig. 2A, Supplementary Data 4). Thus we believe that Hep2 (clone 2B) cells contain at least one full-length Bxv1 provirus. Mean DNA-seq depth of Bxv1 is 2.2 $\times$ , which is lower than that calculated over the entire human genome (6.4 $\times$ ). It is thus possible that the copy that we were able to map is the only one segregating in the cell line. However, we cannot exclude that multiple proviruses are present, which altogether amount to one or less than one provirus per cell (i.e. some cells may be free of integrated Bxv1). The low number of Bxv1 proviruses may be due to the low capacity of the virus to replicate in these cells. Yet, the entire Bxv1 genome is expressed in Hep2 (clone 2B) cells, including the long terminal repeats which are necessary for the virus to replicate and integrate into the host genome, with mean RNA-seq depths varying from 365 $\times$  for the *gag-pro-pol* open reading frame (ORF) to 1955 $\times$  for the *env* ORF (Fig. 2B). However, the level of Bxv1 expression is relatively low, with only 26,985 (or 0.1%) of RNA-seq reads mapping to the virus, which only ranks 21,906th when listing all human transcripts by decreasing number of mapped RNA-seq reads (Supplementary Table 3). This is much lower than the level of expression of Bxv1 measured in the JY B cell line, in which proviral transcripts ranked first in terms of number of aligned reads, being mapped by almost 1,000,000 reads representing 2.6% of all RNA-seq reads<sup>16</sup>.

In the Lin et al.<sup>16</sup> study, 768 RNA-seq reads were found to contain a junction between positions 587 and 5887 of Bxv1 (accession number: JF908815), which corresponds to a splicing event leading to the expression of the *env* ORF. Our search for virus–virus junctions unveiled a nearly identical splicing event, with 757 reads (145 reads when potential PCR duplicates are removed), supporting a junction between positions 587 and 5884 (Fig. 2). The position of the splice donor site matches exactly between our study, that of Lin et al.<sup>16</sup> and the annotation of the N417 MLV (accession number: HQ246218) which is 99.98% identical to Bxv1 JF908815. However, the position of the splice acceptor site differs (5884 in Lin et al.; 5887 here; 5219 in HQ246218 corresponding to 5601 in JF908815), which may be due to context-dependent variations in splicing. To assess whether the presence of Bxv1 in Hep2 (clone 2B) cells could result from mixing with the JY B cell line, we checked for the presence of RNA-seq and/or DNA-seq reads mapping onto the EBV genome, which is known to be present in JY B cells<sup>16</sup>. We did not find any read mapping on the EBV genome. While this suggests that the Bxv1 contamination is not



due to a mix with JY B cells, we cannot exclude that a transient mix occurred between Hep2 (clone 2B) and JY B cells at some point but that JY B cells are no longer detectable. Interestingly, the RNA-seq coverage of the region corresponding to the alternative *env* transcript is three to five times higher than that of the rest of the Bxv1 sequence (Fig. 2B) showing that for some reason, the splicing generating this transcript is markedly favoured in Hep2 (clone 2B) cells. Rather than alternative splicing, the junction we found in RNA-seq reads between positions 587 and 5887 of Bxv1 could be due to transcription of a deleted Bxv1 copy, that would segregate in some cells in addition to a full length copy. To check for the presence of a deleted Bxv1 copy we designed PCR primers on both sides of the deletion. All PCRs performed using those primers were negative, suggesting that the truncated transcripts unlikely result from transcription of a deleted Bxv1 proviral copy. We have not tested whether the higher expression of the *env* ORF translates into an accumulation of the ENV protein, which has been linked to the generation of cytopathic effects in some MLV<sup>42</sup>. That said, we have not observed any cytopathic effect, in agreement with the fact that most MLV do not generate such effects<sup>43</sup>.

In their study of the JY B cell line contaminated by Bxv1, Lin et al.<sup>16</sup> also found evidence of G-to-A editing likely resulting from the activity of the APOBEC3G restriction factor, which induces deamination of cytidine to uridine in single stranded DNA viral intermediates<sup>44</sup>. Specifically, Lin et al.<sup>16</sup> found that 44 out of the 45 SNPs identified in their RNA-seq reads were G-to-A changes, indicating that they likely resulted from transcription of APOBEC3G-edited viral genomes. Here, in agreement with the known absence of APOBEC3G in HeLa cells<sup>45</sup>, the APOBEC3G transcript ranks only 37,052th in the list of human transcripts ordered by the number of mapped reads (Supplementary Table 3). The number of SNPs supported by more than 10 reads and having a frequency > 2% in our RNA-seq data is anyway too low to draw any conclusion on the possible activity of APOBEC3G in Hep2 (clone 2B) cells, but it is worth noting that four of the five SNPs we identified using those criteria in our RNA-seq data are G-to-A changes.

## Conclusion

In this study, we have confirmed that the Hep2 (clone 2B) cell line (Sigma-Aldrich catalog number: 85011412-1VL) is contaminated by HeLa cells based on the characterization of HPV18 integrants and expression patterns. We have further demonstrated that the cell line is also contaminated by a virus nearly identical to the Bxv1 XMLV provirus, the presence of which is likely due to direct contact between the cell line and another contaminated cell line or reagent. Based on sequencing depth, we show that the cell line might contain only one Bxv1 provirus, which we were able to map to the second intron of the *PUS1* gene. While our study does not demonstrate that Bxv1 is able to replicate in Hep2 (clone 2B) cells, it shows that it is expressed at a moderately high level, which may impact various cellular pathways. Thus, the presence of this virus in otherwise HeLa-contaminated Hep2 (clone 2B) cells will have to be taken into consideration in future studies using this cell line to avoid erroneous interpretations of experimental results. Furthermore, this study should also encourage others using this cell line and related ones, such as the (code A) 86030501-1VL, or 85020207-1VL Hep-2C (HeLa derivative) Human Negroid cervix carcinoma (code C), to check for the presence of Bxv1 using the PCR primers we provide in the materials and methods section.

## Data availability

In addition to the Supplementary Fig. 1 and Supplementary Data 1–4 included in this published article, the raw fastq reads generated during the current study are available in the dbGaP repository under accession number phs001944.v1.p1.

Received: 4 October 2019; Accepted: 31 August 2020

Published online: 29 September 2020

## References

- Gartler, S. M. Apparent HeLa cell contamination of human heteroploid cell lines. *Nature* **217**, 750–751 (1968).
- Nelson-Rees, W., Daniels, D. & Flandermeyer, R. Cross-contamination of cells in culture. *Science* **212**, 446–452 (1981).
- Horbach, S. P. J. M. & Halfman, W. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLoS ONE* **12**, e0186281 (2017).
- Nardone, R. M. Eradication of cross-contaminated cell lines: A call for action. *Cell Biol. Toxicol.* **23**, 367–372 (2007).
- Capes-Davis, A. et al. Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int. J. Cancer* **127**, 1–8 (2010).
- Adey, A. et al. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500**, 207–211 (2013).
- Cantalupo, P. G., Katz, J. P. & Pipas, J. M. HeLa nucleic acid contamination in the cancer genome atlas leads to the misidentification of human papillomavirus 18. *J. Virol.* **89**, 4051–4057 (2015).
- Drexler, H. G. & Uphoff, C. C. Mycoplasma contamination of cell cultures: Incidence, sources, effects, detection, elimination, prevention. *Cytotechnology* **39**, 75–90 (2002).
- Merten, O.-W. Virus contaminations of cell cultures—A biotechnological view. *Cytotechnology* **39**, 91–116 (2002).
- Mirjalili, A., Parmoor, E., Moradi Bidhendi, S. & Sarkari, B. Microbial contamination of cell cultures: A 2 years study. *Biologicals* **33**, 81–85 (2005).
- Cao, S. et al. High-throughput RNA sequencing-based virome analysis of 50 lymphoma cell lines from the cancer cell line encyclopedia project. *J. Virol.* **89**, 713–729 (2015).
- Shioda, S. et al. Screening for 15 pathogenic viruses in human cell lines registered at the JCRB Cell Bank: Characterization of in vitro human cells by viral infection. *R. Soc. Open Sci.* **5**, 172472 (2018).
- Uphoff, C. C., Denkmann, S. A., Steube, K. G. & Drexler, H. G. Detection of EBV, HBV, HCV, HIV-1, HTLV-I and -II, and SMRV in human and other primate cell lines. *J. Biomed. Biotechnol.* **2010**, 1–23 (2010).
- Cmarik, J. L., Troxler, J. A., Hanson, C. A., Zhang, X. & Ruscelli, S. K. The human lung adenocarcinoma cell line EKVX produces an infectious xenotropic murine leukemia virus. *Viruses* **3**, 2442–2461 (2011).
- Hempel, H. A., Burns, K. H., De Marzo, A. M. & Sfanos, K. S. Infection of xenotransplanted human cell lines by murine retroviruses: A lesson brought back to light by XMRV. *Front. Oncol.* **3**, 156 (2013).

16. Lin, Z. *et al.* Detection of murine leukemia virus in the Epstein–Barr virus-positive human B-cell line JY, using a computational RNA-Seq-based exogenous agent detection pipeline, PARSES. *J. Virol.* **86**, 2970–2977 (2012).
17. Sfanos, K. S. *et al.* Identification of replication competent murine gammaretroviruses in commonly used prostate cancer cell lines. *PLoS ONE* **6**, e20874 (2011).
18. Takeuchi, Y., McClure, M. O. & Pizzato, M. Identification of gammaretroviruses constitutively released from cell lines used for human immunodeficiency virus research. *J. Virol.* **82**, 12585–12588 (2008).
19. Uphoff, C. C., Lange, S., Denkmann, S. A., Garritsen, H. S. P. & Drexler, H. G. Prevalence and characterization of murine leukemia virus contamination in human cell lines. *PLoS ONE* **10**, e0125622 (2015).
20. Kozak, C. A. The mouse<sup>+</sup> xenotropic<sup>+</sup> gammaretroviruses and their XPR1 receptor. *Retrovirology* **7**, 101 (2010).
21. Kozak, C. Origins of the endogenous and infectious laboratory mouse gammaretroviruses. *Viruses* **7**, 1–26 (2014).
22. Yan, Y. *et al.* Evolution of functional and sequence variants of the mammalian XPR1 receptor for mouse xenotropic gammaretroviruses and the human-derived retrovirus XMRV. *J. Virol.* **84**, 11970–11980 (2010).
23. McALLISTER, R. M. *et al.* C-type virus released from cultured human rhabdomyosarcoma cells. *Nat. New Biol.* **235**, 3–6 (1972).
24. Naseer, A. *et al.* Frequent infection of human cancer xenografts with murine endogenous retroviruses in vivo. *Viruses* **7**, 2014–2029 (2015).
25. Zhang, Y.-A. *et al.* Frequent detection of infectious xenotropic murine leukemia virus (XMLV) in human cultures established from mouse xenografts. *Cancer Biol. Ther.* **12**, 617–628 (2011).
26. Johnson, A. D. & Cohn, C. S. Xenotropic murine leukemia virus-related virus (XMRV) and the safety of the blood supply. *Clin. Microbiol. Rev.* **29**, 749–757 (2016).
27. Paprotka, T. *et al.* Recombinant origin of the retrovirus XMRV. *Science* **333**, 97–101 (2011).
28. Kozak, C. & Rowe, W. Genetic mapping of xenotropic leukemia virus-inducing loci in two mouse strains. *Science* **199**, 1448–1449 (1978).
29. Kirkegaard, J. S. *et al.* Xenotropic retrovirus Bxv1 in human pancreatic  $\beta$  cell lines. *J. Clin. Investig.* **126**, 1109–1113 (2016).
30. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
31. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
32. Gilbert, C. *et al.* Continuous influx of genetic material from host to virus populations. *PLoS Genet.* **12**, e1005838 (2016).
33. Gorphe, P. A comprehensive review of Hep-2 cell line in translational research for laryngeal cancer. *Am. J. Cancer Res.* **9**, 644–649 (2019).
34. Akagi, K. *et al.* Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* **24**, 185–199 (2014).
35. Bouallaga, I., Massicard, S., Yaniv, M. & Thierry, F. An enhanceosome containing the Jun B/Fra-2 heterodimer and the HMG-I(Y) architectural protein controls HPV18 transcription. *EMBO Rep.* **1**, 422–427 (2000).
36. Wang, X., Meyers, C., Wang, H.-K., Chow, L. T. & Zheng, Z.-M. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. *J. Virol.* **85**, 8080–8092 (2011).
37. Oakes, B. *et al.* Contamination of human DNA samples with mouse DNA can lead to false detection of XMRV-like sequences. *Retrovirology* **7**, 109 (2010).
38. Moore, A. E., Sabachewsky, L. & Toolan, H. W. Culture characteristics of four permanent lines of human cancer cells. *Cancer Res.* **15**, 598–602 (1955).
39. Toolan, H. W. Transplantable human neoplasms maintained in cortisone-treated laboratory animals: H.S. No. 1; H.Ep. No. 1; H.Ep. No. 2; H.Ep. No. 3; and H.Emb.Rh. No. 1. *Cancer Res.* **14**, 660–666 (1954).
40. De Rijck, J. *et al.* The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* **5**, 886–894 (2013).
41. Sultana, T., Zamborlini, A., Cristofari, G. & Lesage, P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* **18**, 292–308 (2017).
42. Zhao, X. & Yoshimura, F. K. Expression of murine leukemia virus envelope protein is sufficient for the induction of apoptosis. *J. Virol.* **82**, 2586–2589 (2008).
43. Sliva, K., Erlwein, O., Bittner, A. & Schnierle, B. S. Murine leukemia virus (MLV) replication monitored with fluorescent proteins. *Virol. J.* **1**, 14 (2004).
44. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
45. Kremer, M. *et al.* Vaccinia virus replication is not affected by APOBEC3 family members. *Virol. J.* **3**, 86 (2006).
46. Peccoud, J. *et al.* A survey of virus recombination uncovers canonical features of artificial chimeras generated during deep sequencing library preparation. *G3amp58 GenesGenomesGenetics* **8**, 1129–1138 (2018).
47. Tsai, I. J. *et al.* Summarizing specific profiles in illumina sequencing from whole-genome amplified DNA. *DNA Res.* **21**, 243–254 (2014).

## Acknowledgements

This study was supported by Agence Nationale de la Recherche (ANR-15-CE32-0011-01 TransVir).

## Author contributions

C.G, N.L., A.B.-D., R.C. conceived the study. I.G. performed research. V.L. and C.G. performed research and wrote the first draft of the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-73169-y>.

**Correspondence** and requests for materials should be addressed to C.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020