Transcriptome Profiling of Epstein-Barr Virus Using Nanopore Sequencing

Norbert Moldován
Department of Medical Biology, Faculty of Medicine, University of Szeged
Kálmán Szenthe
Carlsbad Research Organization Ltd
Ferenc Bánáti
RT-Europe Research Center
Ádám Fülöp
Department of Medical Biology, Faculty of Medicine, University of Szeged
Zsolt Csabai
Department of Medical Biology, Faculty of Medicine, University of Szeged
Dóra Tombácz
Department of Medical Biology, Faculty of Medicine, University of Szeged
János Minárovits
Department of Oral Biology and Experimental Dental Research, Faculty of Dentistry, University of Szeged
Zsolt Boldogkői (boldogkoi.zsolt@med.u-szeged.hu)
Department of Medical Biology, Faculty of Medicine, University of Szeged

Research Article

Keywords: Epstein-Barr virus (EBV), Nanopore Sequencing, Transcriptome Profiling

Posted Date: November 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-109629/v1

Transcriptome Profiling of Epstein-Barr Virus Using Nanopore Sequencing

Norbert Moldován[1], Kálmán Szenthe[2], Ferenc Bánáti[3], Ádám Fülöp[1], Zsolt Csabai[1], Dóra Tombácz[1], János Minárovits[4], Zsolt Boldogkői[1*]

[1]Department of Medical Biology, Faculty of Medicine, University of Szeged, Somogyi B. u. 4., 6720 Szeged, Hungary.

[2]Carlsbad Research Organization Ltd., Szabadság u. 2., 9244 Újrónafő, Hungary.

[3]RT-Europe Research Center, Vár tér 2., 9200, Mosonmagyaróvár, Hungary.

[4]Department of Oral Biology and Experimental Dental Research, Faculty of Dentistry, University of Szeged, Tisza Lajos krt. 64., 6720 Szeged, Hungary.

*boldogkoi.zsolt@med.u-szeged.hu

ABSTRACT

Epstein-Barr virus (EBV) is an important human pathogenic gammaherpesvirus with carcinogenic potential. The EBV transcriptome has previously been analyzed using both Illumina-based short read- and Pacific Biosciences RS II-based long-read sequencing technologies. In this work, we use the Oxford Nanopore Technologies MinION platform for the characterization of the EBV transcriptomic architecture. Both amplified and non-amplified cDNA sequencings were applied for the generation of transcription reads, including both oligo-d(T) and random oligonucleotide-primed reverse transcription. EBV transcripts are identified and annotated using the LoRTIA software suite developed in our laboratory. This study detected novel short genes (embedded into longer host genes) containing 5'-truncated in-frame open reading frames (ORFs), which might encode N-terminally truncated proteins. We also detected a number of novel non-coding RNAs and transcript length isoforms encoded by the same genes but differing in their start and/or end sites. This study also reports the discovery of novel splice isoforms, many of which may represent altered coding potential, and of novel Ori-associated RNA molecules. Additionally, novel mono- and polycistronic, as well as complex transcripts have been uncovered. An intricate meshwork of transcriptional overlaps has also been revealed.

Introduction

Epstein-Barr virus (EBV, human gammaherpesvirus 4), is a member of the *Gammaherpesvirinae* subfamily within the family *Herpesviridae*[1]. Spreading predominantly *via* saliva, EBV is highly prevalent in human populations[2]. EBV plays a role in the pathogenesis of Burkitt's lymphoma and other lymphomas, and it is also involved in the development of nasopharyngeal carcinoma and a subset of gastric carcinomas[3,4]. EBV is classified as a Group 1 carcinogenic agent in humans[5].

Primary EBV infection in early childhood is typically mild or symptomless. Later in life, however, it may cause infectious mononucleosis (IM, glandular fever), a lymphoproliferative disease accompanied with pharyngitis, tonsillitis, fever and lymphadenopathy. In the majority of cases, IM is a self-limiting disease, due to a vigorous T cell response directed to EBV-infected, proliferating B cells expressing viral antigens[6]. Although B cells are the primary targets of EBV, the virus also replicates in oropharyngeal epithelial cells. Both B cells and epithelial cells are capable to produce new EBV particles that carry linear, double stranded viral genomes.

After the primary infection, the virus establishes lifelong latency in memory B cells[7]. In latently infected cells, only a limited set of viral genes are expressed from circular, episomal, chromatinized EBV genomes. In addition to memory B cells, EBV-associated lymphoma and carcinoma cells and *in vitro* immortalized lymphoblastoid cell lines (LCLs) can also carry latent EBV genomes. The host cell epigenetic machinery interacts with the viral episomes and the activity of latent EBV promoters is regulated by the epigenetic marks deposited by host enzymes

to the transcriptional control sequences of the viral genome[8]. In latently infected cells, the viral episomes attach to the nuclear matrix at *oriP*, the latent origin of EBV replication and replicate once per cell cycle with the help of the host DNA synthesis machinery[9]. A variety of signals are capable to disrupt latency and induce EBV lytic reactivation both *in vitro* and *in vivo*[10,11]. Induction of EBV lytic replication results in a change from the restricted, latent pattern of EBV gene expression to successive transcription of immediate early, early and late EBV genes.

The immediate gene products designated BZLF1 and BRLF1 are transactivator proteins switching on the transcription of early genes[12,13]. Early gene products include, among others, a core-set of lytic replication proteins shared by *Herpesviridae*[14]. Lytic EBV DNA synthesis occurs in the replication compartments within the host cell nuclei[15]. In contrast to the replication of latent episomes, this unlicensed, exponential amplification of the viral genome is initiated at one of the two copies of *oriLyt*, the lytic replication origin of EBV DNA synthesis[9]. Since transcription of late EBV genes occurs after DNA amplification, it was suggested that during productive replication the immediate early and early EBV genes are transcribed from chromatinized templates, whereas unchromatinized, unmethylated templates are used for the transcription of late genes encoding structural proteins of the virion[16]. EBV late RNA transcription is aided by the viral preinitiation complex[17]. After the synthesis of viral structural proteins, epigenetically naïve, unmethylated, linear dsDNA molecules are packaged into EBV virions[16,18]. These linear genomes undergo circularization, chromatization and epigenetic modification in the newly infected host cells.

Initial studies indicated that all of the viral genes charted on the approximately 170 kb EBV genome are actively transcribed during the lytic cycle[19–22]. Recent studies of the viral transcriptome revealed, however, a more complex pattern of viral gene expression after the disruption of EBV latency in various cell lines. It turned out to be that lytic cycle transcription is bidirectional, and that many newly identified transcribed regions do not code for proteins[23–26]. These data suggested that most probably hundreds of viral long noncoding RNAs (lncRNAs) are generated during productive EBV replication. In addition, novel splicing events further increase the diversity of viral RNAs expressed during productive EBV replication[27].

Earlier studies based on Illumina short-read sequencing (SRS)[17,23,25,28] and Pacific Biosciences (PacBio) RS II long-read sequencing (LRS)[25] identified a large set of EBV transcripts. However, SRS is not optimal for disclosing the transcriptome complexity[29,30] and RS II sequencing has also a limitation for the detection of transcripts falling into a certain size interval[31,32].

In this work, we analyzed the EBV lytic transcriptome using the Oxford Nanopore Technologies (ONT) MinION sequencing platform, which is suitable to produce a complete picture on viral transcripts[33–36], excluding small RNA molecules, such as miRNAs.

Results

Long-read sequencing of the EBV transcriptome

In this work, we analyzed the lytic EBV transcriptome with ONT-based LRS techniques. Oligo(dT)-primed amplified and non-amplified cDNA libraries were generated from eight consecutive lytic time points. However, due to the low coverage, especially at the early time-points kinetic analysis was not feasible. A total of 22,358 non-amplified and 54,271 amplified reads mapped to the viral genome with an average mapped read length of 838,66 bp (σ=701.5) and 1098.43 bp (σ=758.28) respectively. Detailed statistics of the sequencing can be found in Supplementary Table S1. We also generated random hexamer-primed amplified library from the pooled samples, and sequenced them on the MinION platform.

Transcripts with alternative transcript start and end sites

This study identified a total of 213 putative transcriptional start sites (TSSs) using the LoRTIA software suite [https://github.com/zsolt-balazs/LoRTIA] developed in our laboratory (Supplementary Table S2). We used the deepCAGE and PacBio datasets (published by O'Grady et al.[25]) for the validation of our TSSs. A TSS was accepted if it was present in at least two of our samples or in one of our samples and either in the deepCAGE or in the PacBio dataset. This stringent filtering resulted in a total of 183 TSS of which 56 are novel (Fig. 1a). Upstream TATA boxes were identified for 22% of the TSSs at an average distance of -31.31 bp (σ=3.25). The nucleotide composition analysis of these start sites revealed a G-rich initiator region (Fig. 1b).

A total of 58 putative transcription end sites (TESs) were detected using the LoRTIA software suite (Supplementary Table S3). A TES was accepted if it was present in at least two of our samples or in either one of our samples and in the PacBio dataset[25], or the PA-seq dataset (published by Majerciak et al.[26]). This analysis resulted in the detection of 55 TESs, of which 4 is novel (Fig. 1c). Polyadenylation signals (PASs) were identified for 89% of the TESs at an average distance of -23.8 bp (σ=6.14). TESs with a PAS showed an A-rich cleavage site and a G/T-rich downstream element. These sequences are similar to those of the mammalian cleavage and polyadenylation motifs[37]. TESs lacking a PAS showed a AC<u>C</u>TC sequence near the cleavage site (underlined) and a TTATT sequence between +11 and +15 positions (Fig. 1d), the latter being a variant of the termination signal of genes transcribed by RNAPIII[38], a polymerase specialized in transcribing rRNAs and tRNAs. However, it has been described as a terminator for the VA RNA gene of avian adenovirus CELO[39], suggesting the possibility of incidental transcription by RNAPIII for protein coding genes.

We used our validated TSSs, TESs and introns for the annotation of transcript isoforms. This resulted in a total of 306 polyadenylated transcripts (Supplementary Table S4), of which 214 are novel. Some very long transcripts are represented by only a single read, which is below the threshold of detection of LoRTIA. These transcripts are longer than any other overlapping transcripts, but because of their low abundance their TSS is uncertain, hence we denote these as putative transcript isoforms. We detected a total of 25 of these very long putative transcripts (Supplementary Table S5), but we think that a higher data coverage would reveal a much larger population of these transcripts.

A previous LRS study disclosed a large diversity of TSSs and TESs for several EBV genes[25]. Our work revealed 26 novel transcripts with longer and 53 with shorter 5'-UTRs. The deep-CAGE data analysis validated 69% of our longer TSSs and 79% shorter TSSs isoforms.

The 5'-UTRs can regulate translation through their secondary structures[40], upstream AUGs (uAUGs), or upstream ORFs (uORFs)[41,42]. Watanabe and colleagues[43] analyzed the impact of two uORFs upstream of *BGLF3.5* ORF on the translation of *BGLF4,* a protein kinase involved in replication and nuclear egress[44], and found that point mutations in the two uORFs (duORFs) had no effect on protein levels of *BGLF4*. The most abundant transcript of the *BGLF3.5-BGLF4* cluster is *BGLT16*, a bicistronic mRNA including both duAUGs in its 5'-UTR, and an additional uAUG upstream (Fig. 2). We detected two short 5'-UTR isoforms of this transcript (BGLT23 and BGLT25) carrying solely the *BGLF4* gene. This RNA molecule lacks the duORFs mutated by Watanabe and co-workers[43]. Furthermore, BGLT24, a longer 5'-UTR isoform of BGLT16 contained additional uAUGs and uORFs upstream from the modified nucleotides (Fig. 2). We hypothesize that BGLF4 was able to evade translational interruption expected by Watanabe et al.[43] through both its shorter isoforms, or through a complex interplay of uORFs present in the 5'-UTR of its longer isoform.

Our analysis disclosed 6 isoforms with alternative polyadenylation sites. Intriguingly, all of them are located within the same 5 kb region, BZLT44, BZLT45 and BZLT46 are 3'-UTR isoforms of the BZLF2, whereas BELT6, BELT8 and BELT9 are isoforms of the BELT1 transcripts. As a consequence, a 10 nt-long convergent overlap is formed between BZLT44, BELT8, BELT9 and BERT3 (Supplementary Table S4).

Novel monocistronic mRNAs with canonical ORFs

Here we report the detection of 14 novel monocistronic transcripts (Supplementary Table S2). We identified unspliced BNRT10, BHLF1, BORF2 and BGLT18 transcripts of which only spliced versions were previously detected[25,27]. Earlier studies have reported transcripts without ORFs. We also discovered 9 novel monocistronic transcripts with full-length ORFs of which only shorter isoforms with incomplete ORFs (they lack in-frame AUG) have been previously described[25]. Transcripts have not been annotated in the genomic region of *BFRF3*, although studies demonstrated this region as being transcriptionally active[45,46]. We identified BFRT3 fully overlapping the BFRF3 ORF, and ending in a novel terminus (Supplementary Table S2).

Splice junctions and introns

Reverse transcription and PCR may result in gaps in the cDNAs through template-switching (TS) events, which can lead to false intron annotation. The LoRTIA software suite is capable to eliminate these artifacts by detecting the absence of splice junction consensuses or the presence of repeat regions which favors TS. Using the LoRTIA tool kit, we detected a total of 217 putative introns, of which 110 were present in at least 2 samples, 55 of these

being novel. Of those 107 introns present in only a single sample 46 were detected by others[25]. We also accepted these introns, and thus altogether, we obtained 156 introns, which were subjected for further analysis. A canonical GT/AG splice junction consensus was present in every identified intron (Supplementary Table S6).

mRNAs with altered coding potential

Several 5'-truncated transcripts being co-terminal with previously annotated mRNAs lack the canonical ORF, but contain downstream in-frame AUGs, and thus may code for N-terminally truncated proteins[47]. Here we report 15 such RNA molecules, of which the TSSs of 7 were confirmed by the deep-CAGE dataset (Fig. 3a). Besides the alternative transcription initiation within a gene, alternative splicing is also able to produce transcripts with altered coding potential if splicing occurs within the ORF. Forty-four novel splice isoforms and 8 unspliced versions of previously annotated spliced transcripts were detected in this analysis. Twenty-one transcripts contained introns within the ORFs. Among these, 10 frame-shifting, 6 nonsense terminations (through intron retention leading to premature stop codon) (Fig. 3b), a single ORF with deleted amino acids (in-frame deletion) (Fig. 3c) and 4 intergenic terminations (Fig. 3d) were identified. These latter transcripts contain the regular AUG and a new stop codon in an intergenic position.

The coding potential of these transcripts was evaluated using the Coding-Potential Assessment Tool (CPAT) with default settings[48]. First, we evaluated CPAT's performance by running the software on known coding transcript isoforms and isoforms lacking an ORF. This resulted in a sensitivity of 0.87, a specificity of 0.93 and an accuracy of 0.89, suggesting the usability of the default parameters on our dataset. Then we calculated the coding probability of 5'-truncated, alternatively spliced and unspliced transcript isoforms. CPAT analysis led to the result that 9 of the 5'-truncated isoforms, and all - except the ORFs of the 4 splice isoforms with intergenic termination - may have coding potential (Supplementary Table S7). Therefore, these latter transcripts are considered as non-coding RNAs (ncRNAs). To investigate the homology of proteins encoded by alternatively spliced transcripts, we queried the translation of the altered ORFs to the NCBI non-redundant protein database using protein BLAST. BNRT11 and BNRT12 have the first 20 amino acids of the bnrf1 ORF but ends in a stop codon directly after the first splice acceptor position. For BHLF2 splicing results in frame shifting. The first 77.38% of the ORF is identical to the BHLF1 ORF, whereas the amino acids following the splice acceptor position show no similarity to other proteins in the database. The splice acceptor position of BSLT12, BSLT18, BSLT22 and BSLT24 differ from the splice acceptor of the main isoform (BSLT13). This results in altered amino acids following the acceptor position, which do not match any other protein in the database. *BZLT48* and *BZLT50* encode the first 68 amino acids of BZLF2 ORF, whereas the following amino acids and the stop codon is spliced from the transcript. Thus, the altered ORF continues and ends in the second exon of these transcripts, with the amino acids following the acceptor position showing no similarity to proteins in the database. The second splice donor position of BZLT39 and BZLT41 differs from that in the main isoform (BZLF1) resulting in frame shifting, 8 amino acids and a stop codon follows after the corresponding splice acceptor. BZLT40 retains the second, while BZLT43 both of the introns present in BZLF1, both introns having stop codons in-frame. BBLT11 and BBLT13 are unspliced versions of BBLT14 and BBLT12 respectively. The retained intron harbors a stop codon, and the amino acids coded by the intronic nucleotides show no homology to any proteins in NCBI. BILT47 encompasses a spliced version of the LF2 ORF. The resulting splice donor and acceptor positions are in-frame, whereas 81 amino acids of the LF2 ORF are deleted in this spliced ORF. BART15 a splice isoform of the BART transcripts retains the first and the third introns. The first intron encompasses an in-frame stop codon. The resulting altered protein shows partial homology with the first exon of the a73 ORF. The putative proteins of BZLT49, BZLT54 and BZLT51 show no homology to any other entry in NCBI (Supplementary Table S7).

Non-coding transcripts

Transcripts lacking an ORF longer than 10 amino acids were categorized as non-coding. In this part of this study, we detected 3 short ncRNAs (sncRNAs; shorter than 200 bps) and 15 lncRNAs (longer than 200 bps). Twelve of the lncRNAs are 5'-truncated, whereas three of them (BFRT14, BLRT9 and BZLT47) are 3'-truncated isoforms of previously annotated RNAs. BLRT9, a lncRNA starts in the same position as BLRT5 but is terminated 490 bp downstream detected by both our analysis and the Illumina PA-Seq. BLRT9 overlaps the BZTL and BELT region in antisense orientation.

Ori-associated transcripts

Eukaryotic replication origins are largely associated with coding and non-coding transcripts[49,50], Ori-overlapping transcripts were previously demonstrated in alpha-[36,51], beta-[52,53] and gammaherpesviruses[54]. The EBV genome possesses two lytic (OriLyt) and a single latent (OriP) origin of replication. The left OriLyt has been shown to overlap the splice isoforms of BWRT and BCRT, whereas BHLT2 starts within this Ori[25]. The genomic region containing OriP also shows transcriptional activity: several TSSs of various ncRNAs located within the Ori[55]. We detected 5 novel isoforms of Ori-associated RNAs, all of which were initiated within one of the lytic replication origins. BHLF1 and BHLF2 share their TSSs with BHLT2 and are the main transcripts of bhlf1, whereas BHRT15, BHRT16 and BHRT17 are novel 5'-UTR and splice isoforms encoded by the bhrf1 gene (Fig. 4a). The LF3 has previously been shown to start in the right OriLyt, region, but only its ORF and promoter was detected. We annotated LF3 and found 4 other spliced transcripts (RPMS2, RPMS3, RPMS4 and RPMS5) that fully overlap the Ori region, and BILT45 and BIRT21 transcripts of which the 5'-UTR regions overlap the replication origin (Fig. 4b).

Polygenic transcripts

Polygenic transcripts include bi- and polycistronic mRNA molecules and complex transcripts, the later contain at least one gene in an opposite orientation relative to the other gene(s). Polygenic transcripts are abundant in every examined large DNA virus, including herpesviruses[33–36,51,56]. Polycistronic mRNAs have previously been detected in EBV using both SRS[26] and LRS techniques[25]. This study identified 29 novel polycistronic transcripts and demonstrated that basically every lytic gene cluster with the same orientation of ORFs is overlapped by at least one polycistronic transcript isoform. Additionally, three complex transcripts (BBRT18, BGLT29 and BVRT9) containing genes with opposite polarity, were also detected. An overview of the transcripts uncovered in this study is found in Fig. 5.

Transcriptional overlaps

We detected all of the three forms of transcriptional overlaps including divergent (head-to-head), convergent (tail-to-tail) and tandem (parallel, tail-to-head) overlaps between EBV RNAs. These can be formed between transcripts of adjacent genes, like *BDRF1* and *BILF2* or long polygenic and monocistronic transcripts, for example BBRT18, a bicistronic transcript overlapping the isoforms of BBTR16 and the isoforms of BBRT14, both in the same orientation and BBRT18, and the isoforms of BBLT14 in the opposite orientation. Several long spliced transcripts also overlap multiple genes. BDLT32 for example initiates upstream of BDLF2 and overlaps the transcript of 12 genes in the same orientation as BDLF2 and the transcript of 3 genes in reverse orientation. Although transcriptional overlaps represent a common phenomenon in EBV, the intergenic regions between the convergent *BHRF1* and *BHLF1* showed very low level of overlapping transcripts. The intergenic region of *BMRF2* and *BSLF2*/*BMLF1* was found to be devoid of transcriptional activity. However, a higher overall transcript coverage may detect a low-level activity at this region.

Discussion

In this study, we report the profiling of the EBV lytic transcriptome by ONT long-read sequencing platform using non-amplified and amplified cDNA libraries. For the transcript annotation, we used our own dataset together with previous short- and long-read sequencing data[25,26]. A total of 214 novel lytic viral transcripts were detected and 92 previously detected transcript isoforms were confirmed.

A recent study[57] on human adenovirus type 5, a linear dsDNA virus with medium genome size, disclosed a huge plasticity in intron and TES usage. The authors speculate that this flexibility in viral RNA synthesis can lead to selection advantages and thus fuel viral evolution. Previous works on herpesviruses[36,51,56,58,59] including the EBV[25,26] uncovered a great variety of length isoforms, the function of which is still mostly unclear. TSS isoforms through uORFs, uATGs and other cis-acting elements of the 5'-UTRs may play a role in translational regulation[60]. Transcripts with alternative termination can have different turnover times[61,62], localization[63] and altered translation[64].

In this work, we report the detection of novel length variants and splice isoforms that may alter the coding potential of several viral genes. Further proteomic studies are needed to conclude the function of these transcripts, if any.

Our analysis also revealed novel *Ori*-overlapping transcripts. Rennekamp and Lieberman showed in their study[14] that the BHLF1 transcript (overlapping the left *Ori*-lyt) stably binds to its DNA template, and either BHLF1 or the divergent BHRF1 transcript is necessary for the initiation of lytic replication from this *Ori*. We detected the TSS and TES of BHLF1 with base-pair precision, and the existence of a splice isoform of BHLF1 the BHLF2 transcript. We also identified three isoforms of BHRF1, the BHRT15, the BHRT16 and the BHRT17, with a longer than previously detected, *Ori*-overlapping 5'-UTR. The effect of these novel isoforms on the viral replication is yet to be evaluated. This study demonstrated the existence of an extremely complex meshwork of transcriptional overlaps.

Methods

Cells and viruses

Close to saturation, Akata cells were diluted one-to-one with RPMI-1640 medium supplemented with 10% FCS and pens/strep 24 hrs before induction. Cells were washed and resuspended to 10^6 cell/ml in RPMI solution supplemented with Goat anti-human IgG (Jackson, 109-001-003, 17μg/ml final concentration), or in normal RPMI serving as controls in 7-7 T25 cell culture flasks[23]. 100 μl cell suspensions were aspirated at time points of 10 min, 90 min, 4, 12, 24, 48 and 72 hrs after resuspension for RNA isolation to verify the success of induction of Epstein-Barr Virus transcription. Applying real-time PCR the activity of the BZLF and GP350 genes were monitored normalized to reference genes. The remaining cells were pelleted and stored at -70°C.

Assessment of the lytic induction

Total RNA was isolated from 100 μl cell suspensions with Direct-zol method (Zymo) as recommended by the manufacturer. DNase I treatment (Thermo, EN0525) was performed subsequently according to the manufacturer. RNA samples were amplified in a one-step reaction using oligonucleotides specific for BZLF, GP350 transcripts and ACTB and GADPH as reference genes (Table 1) with the SYBR-Green based approach (Bioline, SensiFAST™ SYBR® No-ROX One-Step Kit, BIO-72001) as recommended in the manual as follows: reverse transcription: 50°C, 10 min; Initial denaturation: 95°C 2 min, 45 cycles of 95°C, 10 sec and 60°C, 30 sec. Amplification specificity was checked with melting curve analysis.

Table 1. Primers used for the assessment of lytic induction.

Primer name	Sequence
BZLF1 F	CCCAAACTCGACTTCTGAAGATGTA
BZLF1 R	TGATAGACTCTGGTAGCTTGGTCAA
gp350 F	AGAATCTGGGCTGGGACGTT
gp350 R	ACATGGAGCCCGGACAAGT
ACTB fw	GGCGGCACCACCATGTACCCT
ACTB rv	AGGGGCCGGACTCGTCATACT
GAPDH fw	GGAAGGTGAAGGTCGGAGTCA
GAPDH rv	ATGGGTGGAATCATATTGGAACA

RNA isolation for sequencing

Total RNA was purified from the cells using the NucleoSpin RNA Kit (Macherey-Nagel). Total RNA samples were split in two. Polyadenylated RNAs were isolated from half of the total RNA samples using the Oligotex mRNA Mini Kit (Qiagen). Ribodepletion, was carried out to remove ribosomal RNA from the other half of total RNAs using Epicentre Ribo-Zero Magnetic Kit. The concentrations of RNA samples were determined using Qubit 4 (Thermo Fisher Scientific). The RNA BR Assay Kit (Thermo Fisher Scientific) was used for the quantification of total RNAs while the Qubit RNA HS Assay (Thermo Fisher Scientific) Kit was applied for the measurement of polyadenylated and ribodepleted samples. The RNA quality was measured with a TapeStation 4150 (Agilent).

Amplified cDNA library preparation

Amplified cDNA libraries were prepared from the purified polyA(+) RNAs using ONT Ligation Kit 1D (SQK-LSK109). Briefly, 50 ng of polyA selected RNA was reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) and (VN)T_{20} (oligo dT) primers (from the ONT kit). The cDNA samples were amplified using Kapa HiFi PCR Kit (Kapa Biosystems), followed by end repair treatment [NEBNext End repair/dA-tailing Module (New England Biolabs)]. The libraries were barcoded using 1D PCR Barcoding (96) Kit (ONT) following the manufacturer's instructions. Between each step the samples were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter). The concentration of the libraries was determined by Qubit 4 Fluorometer. Barcoded libraries were pooled in equimolar ratios and 200 fmol of the pooled sample was loaded on a MinION Flow Cell.
Amplified cDNA library was also generated from 50 ng ribodepleted RNA and custom-made random primers. The consecutive steps were the same as described above.

Non-amplified cDNA library preparation

ONT Direct cDNA Sequencing Kit (SQK-DCS109) was used for the generation of amplification-free libraries. In short, 100 ng polyA-selected RNA sample was used for the synthesis of the first cDNA strand using Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) RNase Cocktail Enzyme Mix (Thermo Fisher Scientific) was used for the removal of RNAs from the single stranded cDNA molecules. The synthesis of the second cDNA strand was performed using LongAmp Taq Master Mix (New England Biolabs). cDNA ends were repaired using NEBNext Ultra II End Repair/dA-Tailing Module. The cDNA ends were repaired using NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs). Libraries were barcoded using ONT Native Barcoding Expansion Kit (EXP-NBD104), then the ligation of the sequencing adapter was carried out using NEB Quick T4 DNA Ligase. All conditions were set according to the SQK-DCS109 manufacturer's protocol.

Pre-processing and data analysis

MinION data were base-called and demultiplexed using Guppy base caller v. 3.3.3. with *–qscore_filtering* turned on. Reads with a Q-score larger than 7 were mapped to the circularized viral genome (NCBI nucleotide accession: KC207813.1) using the Minimap2 software[65]. Adapter sequences and poly(A) tails were preserved on reads to determine 5' and 3' ends and the orientation of the transcripts. Previously published[25,26] DeepCAGE, PA-seq and PacBio RSII data were retrieved for TSS, TES and transcript validation.

Isoform annotation

Transcript isoform detection and annotation was carried out using the LoRTIA software suite v.0.9.9 [https://github.com/zsolt-balazs/LoRTIA] as follows: (1) Artefacts resulting from false priming or partial RT or PCR were removed by searching non-trimmed read ends for the sequencing adapters for the TSS or a homopolymer A sequence for the TES. The first nucleotide not aligning with the adapter or the homopolymer A was denoted as possible TSS or TES. Any other read ends were excluded from further analysis; (2) Random start and end positions caused by RNA degradation were further filtered by testing the putative TSSs and TESs against the Poison distribution, with the significance corrected by the Bonferroni method[66]. Features failing to qualify as local maxima, or being present in less than 1‰ of the coverage were eliminated from the analysis; (3) Gaps were denoted as putative introns, if they have one of the three most frequent consensus sequence (*GT/AG, GC/AG, AT/AC*) and if they are more abundant than 1‰ compared to the local coverage. Putative introns flanked by tandem repeat regions were removed from the analysis as possible template-switching artefacts.

Putative TSSs and TESs were accepted as real if presented in either two of our samples, or for TSSs in one of our samples and one in either deepCAGE or PacBio results, for TESs in one of our samples and either PA-Seq or the PacBio results.

Likewise, putative introns were accepted as real if they were present in two of our samples, or in one of our samples and the PacBio results. Accepted TSSs, TESs and introns were then assembled into transcripts using the Transcript_Annotator software of the LoRTIA toolkit. Very long unique or low-abundance reads which could not be detected using LoRTIA were evaluated and annotated manually. These reads were also accepted as putative transcript isoforms if they were longer than any other overlapping RNA molecule.

Transcript nomenclature

We used the conventional terminology for naming the EBV transcriptome[25]. Novel transcript isoforms were named after the most abundant previously annotated transcript of a gene.

Coding potential estimation

In order to estimate coding potential of the transcripts with previously undetected ORFs, we extracted the transcript sequences from the reference genome and used the Coding-Potential Assessment Tool (CPAT)[48] with default settings.

References

1. Davison, A. J. *et al.* The order Herpesvirales. *Archives of Virology* vol. 154 171–177 (2009).

2. Rickinson A, K. Epstein-Barr Virus. in *Fields Virology* (eds. D., K. & P., H.) 2655–2700 (Lippincott, Williams and Wilkins, 2007).

3. Young, L. S., Yap, L. F. & Murray, P. G. Epstein-Barr virus: More than 50 years old and still providing surprises. *Nature Reviews Cancer* vol. 16 789–802 (2016).

4. Shannon-Lowe, C. & Rickinson, A. The Global Landscape of EBV-Associated Tumors. *Frontiers in Oncology* vol. 9 713 (2019).

5. de Martel, C., Georges, D., Bray, F., Ferlay, J. & Clifford, G. M. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. Heal.* **8**, e180–e190 (2020).

6. Meckiff, B. J. *et al.* Primary EBV Infection Induces an Acute Wave of Activated Antigen-Specific Cytotoxic CD4 + T Cells. *J. Immunol.* **203**, 1276–1287 (2019).

7. Thorley-Lawson, D. A. EBV persistence-introducing the virus. in *Epstein Barr Virus* vol. 1 151–209 (Springer International Publishing, 2015).

8. Takacs, M. *et al.* Epigenetic regulation of latent Epstein-Barr virus promoters. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1799 228–235 (2010).

9. Hammerschmidt, W. & Sugden, B. Replication of Epstein-Barr viral DNA. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).

10. Li, H. *et al.* Therapies based on targeting Epstein-Barr virus lytic replication for EBV-associated malignancies. *Cancer Science* vol. 109 2101–2108 (2018).

11. Kenney, S. C. & Mertz, J. E. Regulation of the latent-lytic switch in Epstein-Barr virus. *Seminars in Cancer Biology* vol. 26 60–68 (2014).

12. Liu, P. & Speck, S. H. Synergistic autoactivation of the Epstein-Barr virus immediate-early BRLF1 promoter by Rta and Zta. *Virology* **310**, 199–206 (2003).

13. Schaeffner, M. *et al.* BZLF1 interacts with chromatin remodelers promoting escape from latent infections with EBV. *Life Sci. Alliance* **2**, (2019).

14. Rennekamp, A. J. & Lieberman, P. M. Initiation of Epstein-Barr Virus Lytic Replication Requires Transcription and the Formation of a Stable RNA-DNA Hybrid Molecule at OriLyt. *J. Virol.* **85**, 2837–2850 (2011).

15. Nagaraju, T., Sugden, A. U. & Sugden, B. Four-dimensional analyses show that replication compartments are clonal factories in which Epstein–Barr viral DNA amplification is coordinated. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24630–24638 (2019).

16. Chakravorty, A., Sugden, B. & Johannsen, E. C. An Epigenetic Journey: Epstein-Barr Virus Transcribes Chromatinized and Subsequently Unchromatinized Templates during Its Lytic Cycle. *J. Virol.* **93**, 2247–2265 (2019).

17. Djavadian, R., Hayes, M. & Johannsen, E. CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. *PLOS Pathog.* **14**, e1007114 (2018).

18. Woellmer, A. & Hammerschmidt, W. Epstein-Barr virus and host cell methylation: Regulation of latency, replication and virus reactivation. *Current Opinion in Virology* vol. 3 260–265 (2013).

19. Yuan, J., Cahir-McFarland, E., Zhao, B. & Kieff, E. Virus and Cell RNAs Expressed during Epstein-Barr Virus Replication. *J. Virol.* **80**, 2548–2565 (2006).

20. Dresang, L. R. *et al.* Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: Insights on the detection and discovery of viral genes. *BMC Genomics* **12**, 625 (2011).

21. Arvey, A. *et al.* An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe* **12**, 233–245 (2012).

22. Ersing, I. *et al.* A Temporal Proteomic Map of Epstein-Barr Virus Lytic Replication in B Cells. *Cell Rep.* **19**, 1479–1493 (2017).

23. O'Grady, T. *et al.* Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. *J. Virol.* **88**, 1604–16 (2014).

24. Cao, S. *et al.* New Noncoding Lytic Transcripts Derived from the Epstein-Barr Virus Latency Origin of Replication, oriP, Are Hyperedited, Bind the Paraspeckle Protein, NONO/p54nrb, and Support Viral Lytic Transcription. *J. Virol.* **89**, 7120–7132 (2015).

25. O'Grady T, Wang X, Höner Zu Bentrup K, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Res. 2016 Oct 14;44(18): e145.

26. Majerciak, V., Yang, W., Zheng, J., Zhu, J. & Zheng, Z.-M. A Genome-Wide Epstein-Barr Virus Polyadenylation Map and Its Antisense RNA to EBNA. *J. Virol.* **93**, (2018).

27. Concha, M. *et al.* Identification of New Viral Genes and Transcript Isoforms during Epstein-Barr Virus Reactivation using RNA-Seq. *J. Virol.* **86**, 1458–1467 (2012).

28. Peng, R. J. *et al.* Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. *Leukemia* **33**, 1451–1462 (2019).

29. Liu, L. *et al.* Comparison of Next-Generation Sequencing Systems. *J. Biomed. Biotechnol.* **2012**, 1–11 (2012).

30. Steijger, T. *et al.* Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* **10**, 1177–1184 (2013).

31. Balázs, Z. *et al.* Long-Read Sequencing of Human Cytomegalovirus Transcriptome Reveals RNA Isoforms Carrying Distinct Coding Potentials. *Sci. Rep.* **7**, 15989 (2017).

32. *Template Preparation and Sequencing Guide*. www.pacificbiosciences.com (2010).

33. Tombácz, D. *et al.* Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. *Front. Genet.* **10**, 834 (2019).

34. Moldován, N. *et al.* Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci. Rep.* **8**, 8604 (2018).

35. Moldován, N. *et al.* Multi-platform analysis reveals a complex transcriptome architecture of a circovirus. *Virus Res.* **237**, 37–46 (2017).

36. Prazsák, I. *et al.* Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* **19**, 873 (2018).

37. Tian, B. & Graber, J. H. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* **3**, 385–96 (2012).

38. Bogenhagen, D. F. & Brown, D. D. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. *Cell* **24**, 261–270 (1981).

39. Gunnery, S., Ma, Y. & Mathews, M. B. Termination sequence requirements vary among genes transcribed by RNA polymerase III. *J. Mol. Biol.* **286**, 745–757 (1999).

40. Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* **19**, 158–174 (2018).

41. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–12 (2009).

42. Kronstad, L. M., Brulois, K. F., Jung, J. U. & Glaunsinger, B. A. Dual short upstream open reading frames control translation of a herpesviral polycistronic mRNA. *PLoS Pathog.* **9**, e1003156 (2013).

43. Watanabe, T. *et al.* Roles of Epstein-Barr virus BGLF3.5 gene and two upstream open reading frames in lytic viral replication in HEK293 cells. *Virology* **483**, 44–53 (2015).

44. Gershburg, E., Raffa, S., Torrisi, M. R. & Pagano, J. S. Epstein-Barr Virus-Encoded Protein Kinase (BGLF4) Is Involved in Production of Infectious Virus. *J. Virol.* **81**, 5407–5412 (2007).

45. Han, Z., Verma, D., Hilscher, C., Dittmer, D. P. & Swaminathan, S. General and Target-Specific RNA Binding Properties of Epstein-Barr Virus SM Posttranscriptional Regulatory Protein. *J. Virol.* **83**, 11635–11644 (2009).

46. Batisse, J., Manet, E., Middeldorp, J., Sergeant, A. & Gruffat, H. Epstein-Barr Virus mRNA Export Factor EB2 Is Essential for Intranuclear Capsid Assembly and Production of gp350. *J. Virol.* **79**, 14102–14111 (2005).

47. Crofts, L. A., Hancock, M. S., Morrison, N. A. & Eisman, J. A. Multiple promoters direct the tissue-specific expression of novel N-terminal variant human vitamin D receptor gene transcripts. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10529–10534 (1998).

48. Wang, L. *et al.* CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, 1–7 (2013).

49. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genet.* **9**, e1003569 (2013).

50. Sequeira-Mendes, J. *et al.* Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet.* **5**, 1000446 (2009).

51. Tombácz, D. *et al.* Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLoS One* **11**, e0162868 (2016).

52. Tai-Schmiedel, J. *et al.* The virally encoded long non-coding RNA4.9 is controlling viral DNA replication. in *International Herpesvirus Workshop 2018* 2.32 (University of British Columbia, 2018).

53. Gatherer, D. *et al.* High-resolution human cytomegalovirus transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19755–60 (2011).

54. Wang, Y., Tang, Q., Maul, G. G. & Yuan, Y. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator. *J. Virol.* **80**, 12171–86 (2006).

55. Cao, S. *et al.* New Noncoding Lytic Transcripts Derived from the Epstein-Barr Virus Latency Origin of Replication, oriP, Are Hyperedited, Bind the Paraspeckle Protein, NONO/p54nrb, and Support Viral Lytic Transcription . *J. Virol.* **89**, 7120–7132 (2015).

56. Balázs, Z., Tombácz, D., Szűcs, A., Snyder, M. & Boldogkői, Z. Long-read sequencing of the human cytomegalovirus transcriptome with the Pacific Biosciences RSII platform. *Sci. Data* **4**, 170194 (2017).

57. Donovan-Banfield, I., Turnell, A. S., Hiscox, J. A., Leppard, K. N. & Matthews, D. A. Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution. *Commun. Biol.* **3**, 1–14 (2020).

58. Moldován, N. *et al.* Multi-Platform Sequencing Approach Reveals a Novel Transcriptome Profile in Pseudorabies Virus. *Front. Microbiol.* **8**, 2708 (2018).

59. Depledge, D. P. *et al.* Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* **10**, 754 (2019).

60. Geballe, A. P. & Mocarski, E. S. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. *J. Virol.* **62**, 3334–40 (1988).

61. Mayr, C. & Bartel, D. P. Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**, 673–684 (2009).

62. Pereira, L. A., Munita, R., González, M. P. & Andrés, M. E. Long 3'UTR of Nurr1 mRNAs is targeted by miRNAs in mesencephalic dopamine neurons. *PLoS One* **12**, e0188177 (2017).

63. Macdonald, P. M. & Struhl, G. Cis- acting sequences responsible for anterior localization of bicoid mRNA in Drosophila embryos. *Nature* **336**, 595–598 (1988).

64. Martin, K. C. & Ephrussi, A. mRNA Localization: Gene Expression in the Spatial Dimension. *Cell* vol. 136 719–730 (2009).

65. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

66. Haynes, W. Bonferroni Correction. in *Encyclopedia of Systems Biology* 154–154 (Springer New York, 2013). doi:10.1007/978-1-4419-9863-7_1213.

Acknowledgements

This study was supported by National Research, Development and Innovation Office (Hungary) grants K 128247 to ZBo and FK 128252 to DT. The publications cost was covered by the University of Szeged Open Access Fund (Grant no: 5060). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions statement

N.M., and Á.F. carried out the bioinformatic analysis of the viral transcripts. Z.C. and D.T. prepared the ONT MinION libraries, carried out ONT sequencing and participated in the analysis. Z.C. and K.S. isolated RNA. K.S. and F.B. maintained cell cultures and performed and tested the reactivation of EBV. N.M and J.M. drafted the manuscript. Z.B. integrated the data and wrote the final version of the manuscript. All authors reviewed and approved the final version of the manuscript.

Additional information

Accession codes: The LoRTIA software suite is available on GitHub: https://github.com/zsolt-balazs/LoRTIA. Our in-house scripts used to generate the descriptive statistics of reads and transcripts, to analyze promoters and to detect transcript isoforms are also available on GitHub: https://github.com/moldovannorbert/seqtools. The sequencing datasets generated during this study are available at the European Nucleotide Archive's SRA database under the accession PRJEB38992 (https://www.ebi.ac.uk/ena/browser/view/PRJEB38992). The transcript annotations are available in the Supplementary Dataset 1.
Competing interests: The authors declare no competing interests.
Ethics declaration: Neither human nor animal experiments were applied in this study.

Figure legends

Fig. 1. TSS and TES proportions and sequence motifs. The proportion of previously annotated, novel and putative (a) TSSs and (c) TESs with the extracted slices represent the proportion of TSSs or TESs with TATA box or PAS respectively, while the rest of the pie chart represent the proportion of TSSs or TESs without a TATA box or PAS. The sequence surrounding the (b) TSS shows a G-rich initiator region for TSSs with a TATA box. TSSs lacking a TATA box have a G-rich +1 and +2 position. (d) TESs with a PAS have a canonical A-rich cleavage site and a canonical GU-rich downstream element (DSE), while those without a PAS showed a C-rich cleavage with no recognizable DSE.

Fig. 2. Transcripts overlapping BGLF4 and BGLF3.5 ORFs and their upstream open reading frames (uORFs). DuORFs are upstream ORFs containing a point-mutation[43]. The gray histogram represents the deep-CAGE data generated by O'Grady and colleauges[25], while the red arrow heads show the concordance of this data with the TSSs of transcript isoforms.

Fig. 3. Transcripts with altered coding potential. (a) Transcript isoforms with TSSs located within the canonical ORFs of the host genes in many cases contain in-frame truncated ORFs. The gray histogram illustrates the deep-CAGE data generated by O'Grady and colleagues.[25]. (b) Alternative splicing of the BZLF1 transcripts results in nonsense termination of the ORFs highlighted by the orange wide rectangles. (c) Splicing of the BILT47 transcript does not cause frameshift but an ORF with deleted nucleotides. (d) Splicing of the unspliced BZLT10 leads to the deletion of a large portion of the ORF and results in an intergenic stop codon (first exon shown by green, while the second exon shown by the orange wide rectangles).

Fig. 4. Novel *Ori*-overlapping RNAs. (a) Three novel length and two novel splice transcript isoforms are initiated in the left lytic *Ori* region. The deep-CAGE data from O'Grady et al.[25] is represented by the histograms with gray reads mapping to the positive, while red are reads mapping to the negative strand. Light blue arrow heads show matching between the deep-CAGE data and our TSS data. (b) Two novel length and five novel splice isoforms initiated or overlapping the right lytic *Ori* region.

Fig. 5. The lytic transcriptome of EBV. Previously detected[25] and novel transcript isoforms are shown on the genome of strain Akata (NCBI accession: KC207813.1). The first 45kbp region of the genome is duplicated on the end to show transcripts overlapping the genomic junction.

Figures

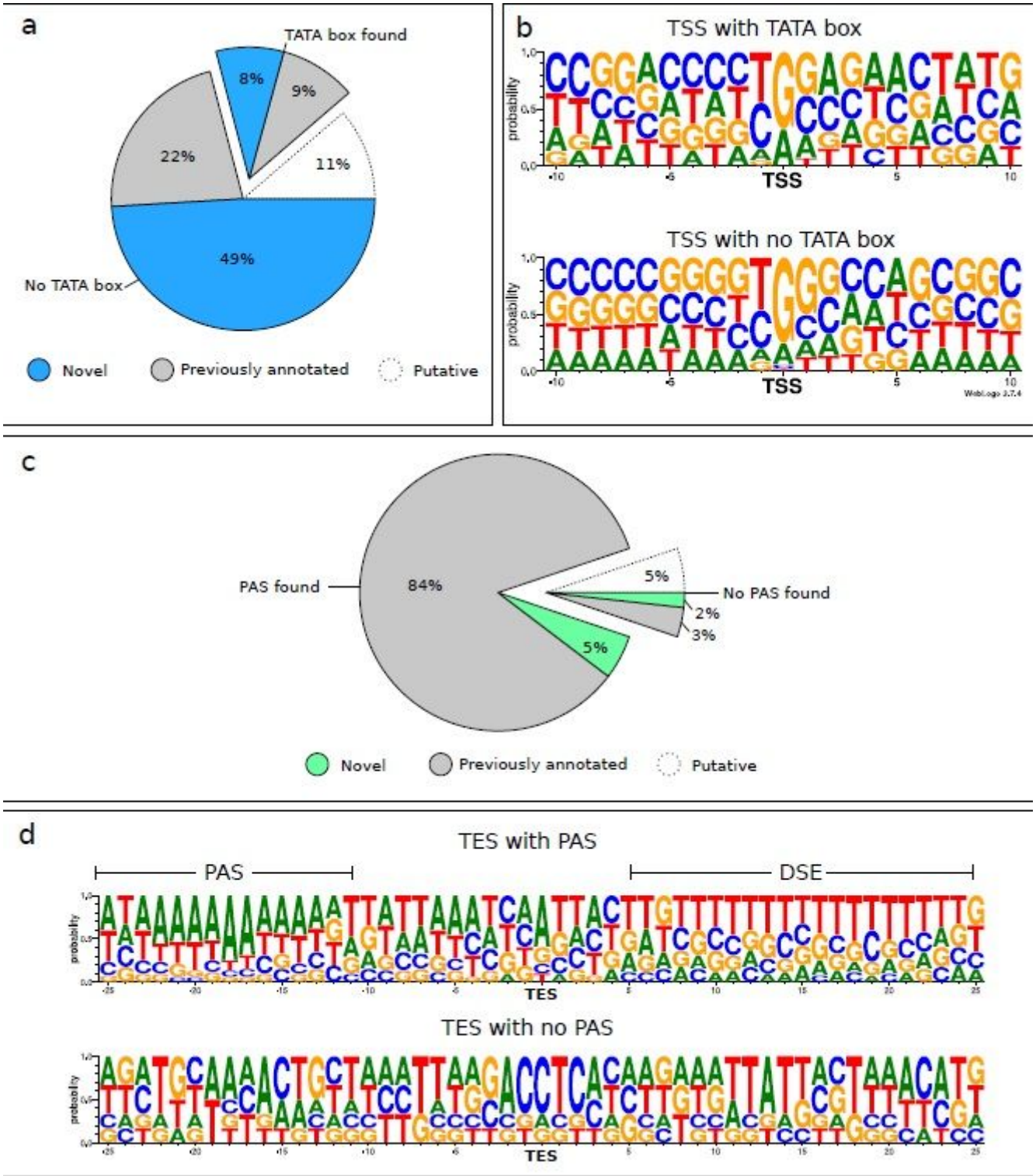

Figure 1

TSS and TES proportions and sequence motifs. The proportion of previously annotated, novel and putative (a) TSSs and (c) TESs with the extracted slices represent the proportion of TSSs or TESs with TATA box or PAS respectively, while the rest of the pie chart represent the proportion of TSSs or TESs

without a TATA box or PAS. The sequence surrounding the (b) TSS shows a G-rich initiator region for TSSs with a TATA box. TSSs lacking a TATA box have a G-rich +1 and +2 position. (d) TESs with a PAS have a canonical A-rich cleavage site and a canonical GU-rich downstream element (DSE), while those without a PAS showed a C-rich cleavage with no recognizable DSE.

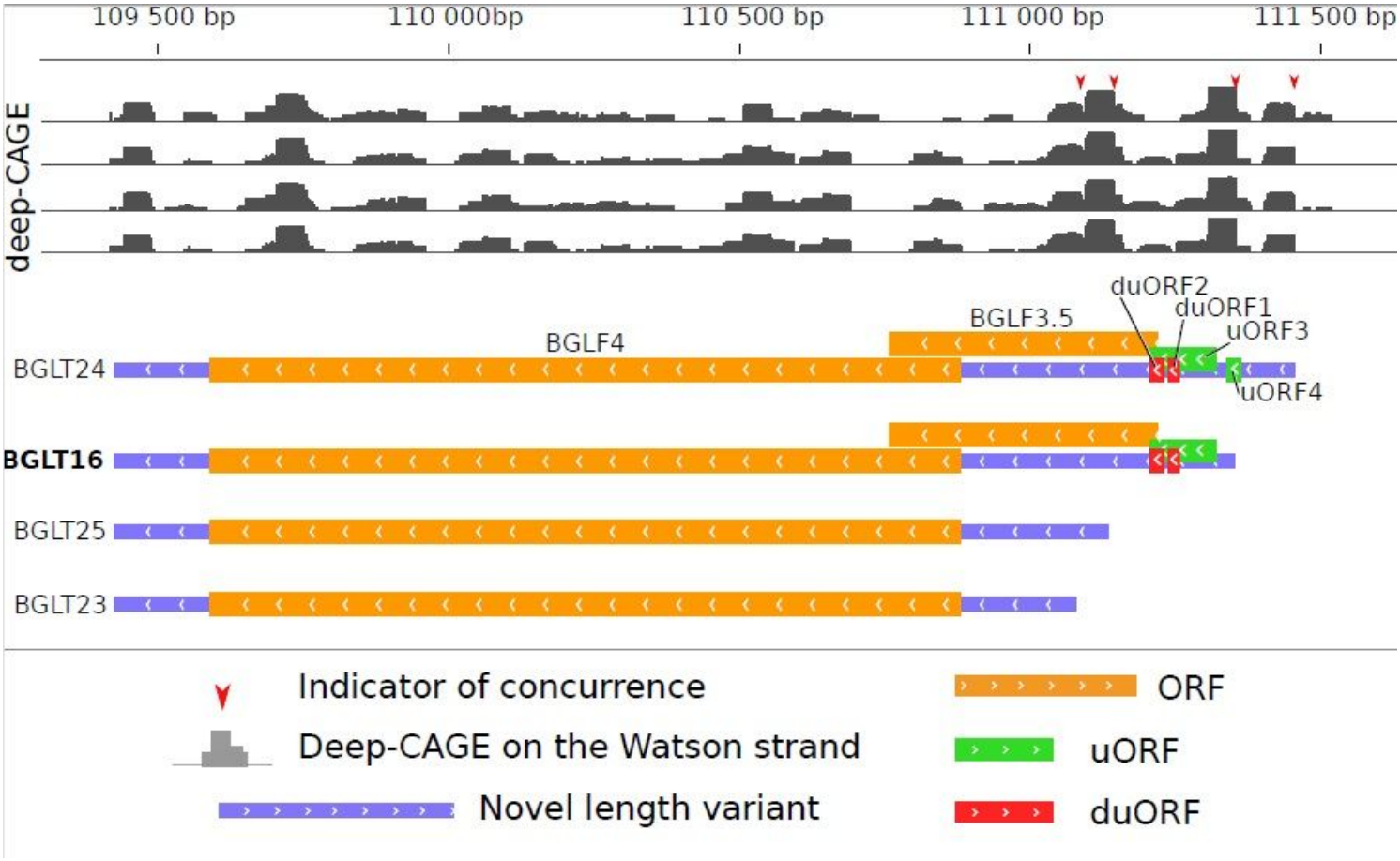

Figure 2

Transcripts overlapping BGLF4 and BGLF3.5 ORFs and their upstream open reading frames (uORFs). DuORFs are upstream ORFs containing a point-mutation43. The gray histogram represents the deep-CAGE data generated by O'Grady and colleauges25, while the red arrow heads show the concordance of this data with the TSSs of transcript isoforms.

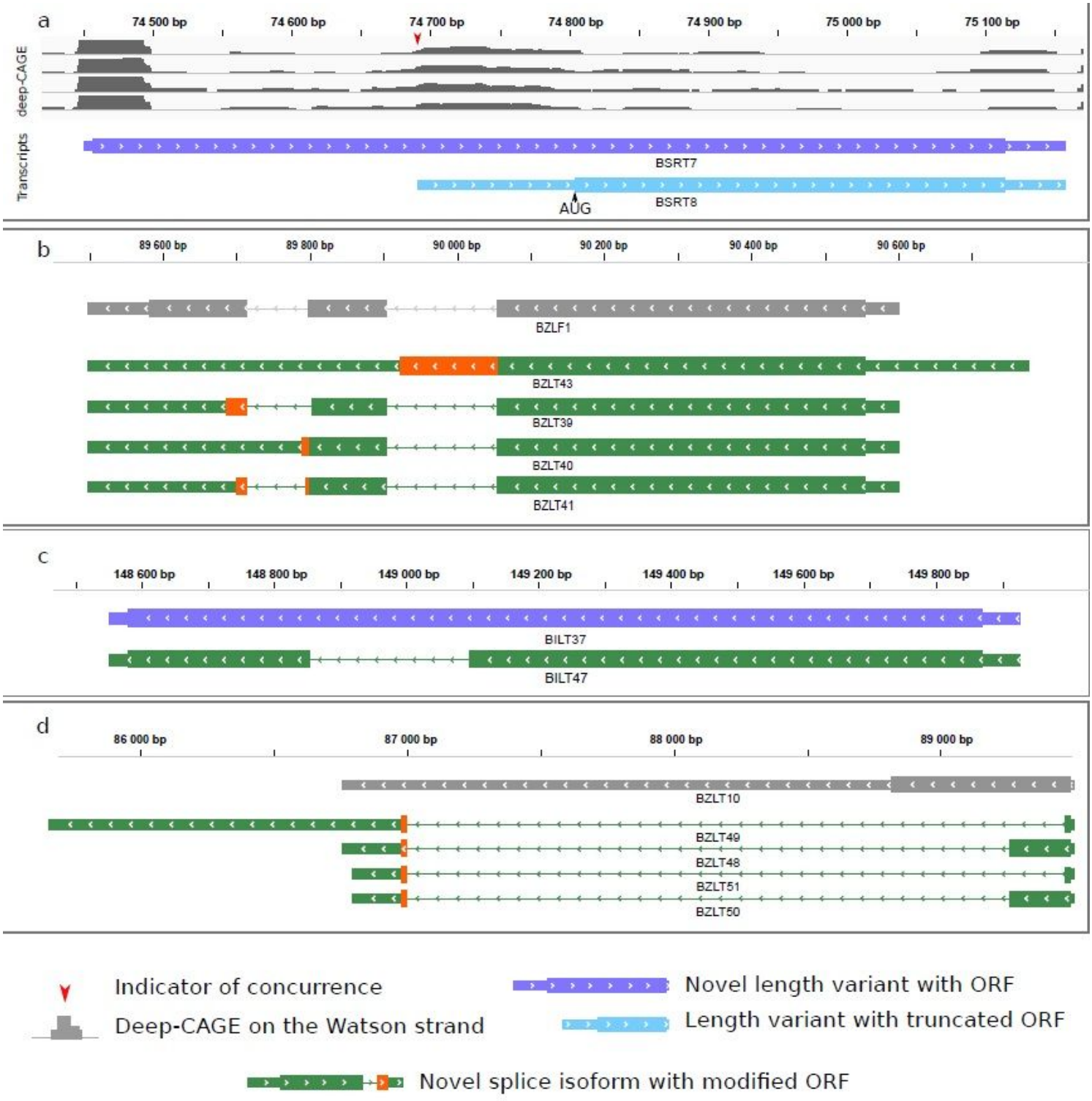

Figure 3

Transcripts with altered coding potential. (a) Transcript isoforms with TSSs located within the canonical ORFs of the host genes in many cases contain in-frame truncated ORFs. The gray histogram illustrates the deep-CAGE data generated by O'Grady and colleagues.25. (b) Alternative splicing of the BZLF1 transcripts results in nonsense termination of the ORFs highlighted by the orange wide rectangles. (c) Splicing of the BILT47 transcript does not cause frameshift but an ORF with deleted nucleotides. (d)

Splicing of the unspliced BZLT10 leads to the deletion of a large portion of the ORF and results in an intergenic stop codon (first exon shown by green, while the second exon shown by the orange wide rectangles).

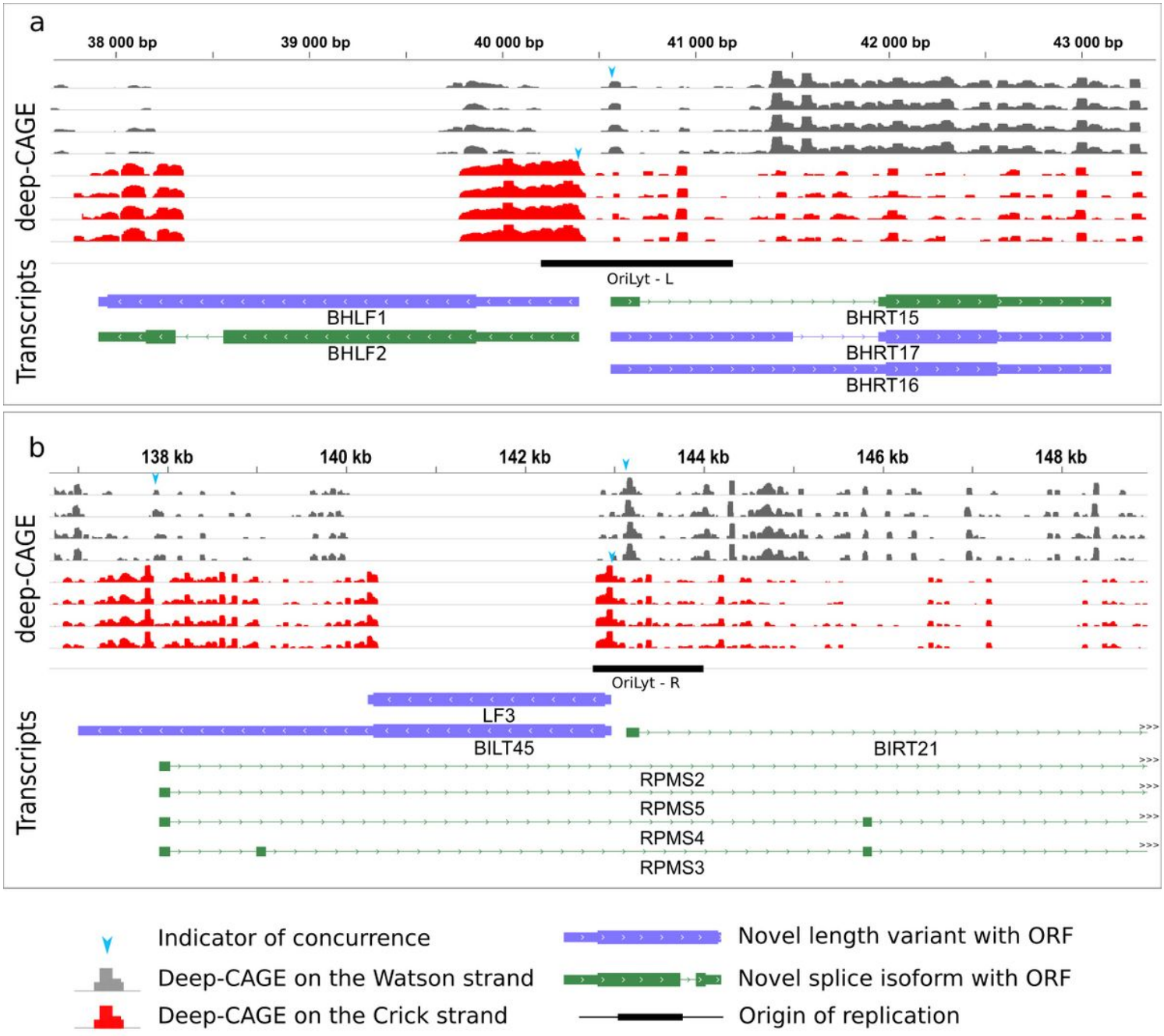

Figure 4

Novel Ori-overlapping RNAs. (a) Three novel length and two novel splice transcript isoforms are initiated in the left lytic Ori region. The deep-CAGE data from O'Grady et al.25 is represented by the histograms with gray reads mapping to the positive, while red are reads mapping to the negative strand. Light blue arrow heads show matching between the deep-CAGE data and our TSS data. (b) Two novel length and five novel splice isoforms initiated or overlapping the right lytic Ori region.

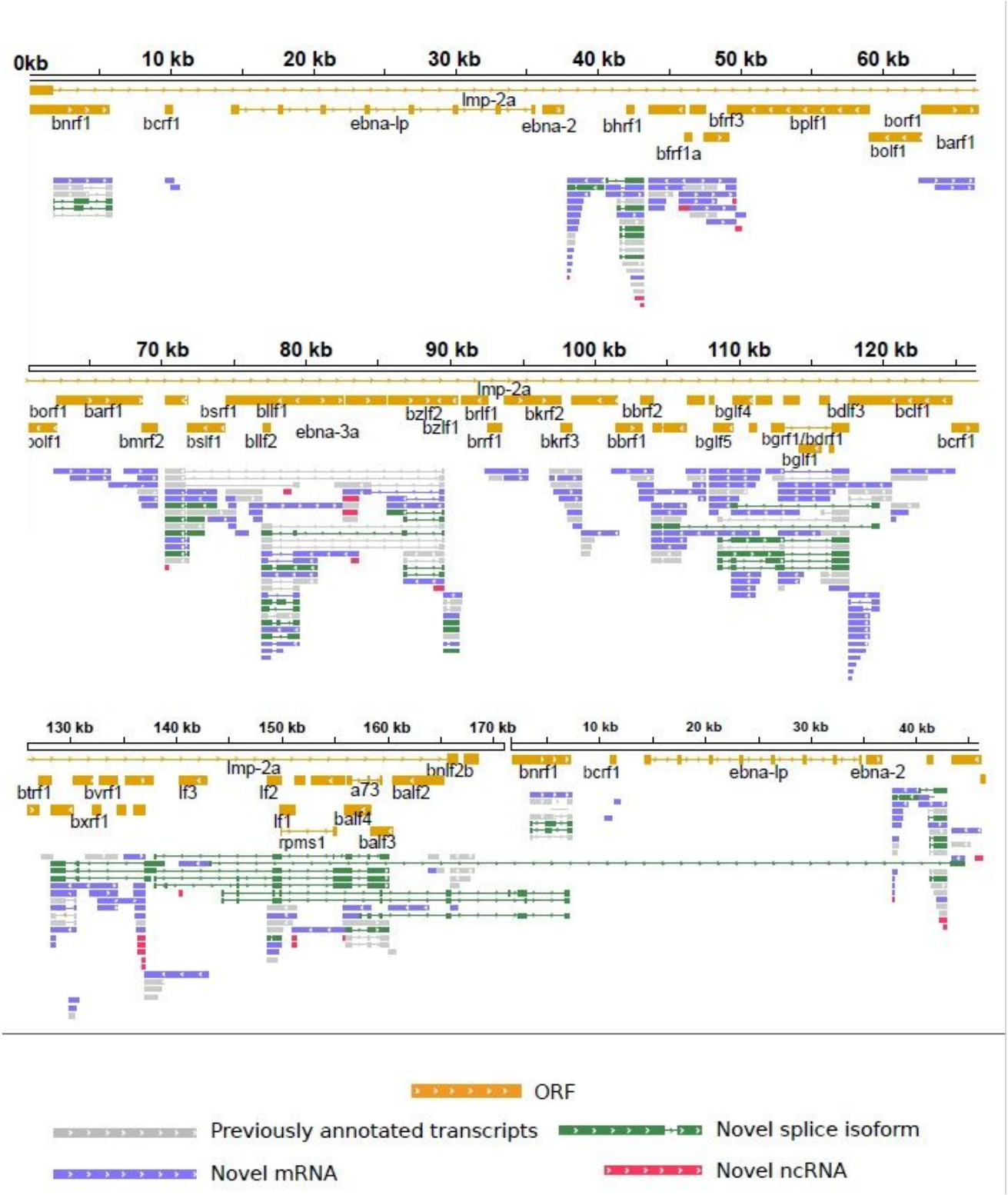

Figure 5

The lytic transcriptome of EBV. Previously detected25 and novel transcript isoforms are shown on the genome of strain Akata (NCBI accession: KC207813.1). The first 45kbp region of the genome is duplicated on the end to show transcripts overlapping the genomic junction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation.docx
- Supplementarytable1.xlsx
- Supplementarytable2.xlsx
- Supplementarytable3.xlsx
- Supplementarytable4.xlsx
- Supplementarytable5.xlsx
- Supplementarytable6.xlsx
- Supplementarytable7.xlsx