Transcriptomic analysis of a human and an insect DNA virus using an integrated sequencing approach

Ph.D. Thesis Booklet

Ádám Fülöp

Department of Medical Biology

Doctoral School of Interdisciplinary Medicine

Faculty of Medicine

University of Szeged

Supervisor: Prof. Dr. Zsolt Boldogkői Ph.D., DSc

Szeged

- 2024 -

Publications directly related to the subject of the thesis

Fülöp Á, Torma G, Moldován N, Szenthe K, Bánáti F, Almsarrhad IAA, Csabai Z, Tombácz D, Minárovits J, Boldogkői Z. Integrative profiling of Epstein-Barr virus transcriptome using a multiplatform approach. Virol J. 2022 Jan 6;19(1):7. doi: 10.1186/s12985-021-01734-6. **IF:4.8**

Torma G, Tombácz D, Moldován N, **Fülöp Á**, Prazsák I, Csabai Z, Snyder M, Boldogkői Z. Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. Sci Rep. 2022 Jan 25;12(1):1291. doi: 10.1038/s41598-022-05457-8. **IF:4.6**

Introduction

All the RNA molecules expressed at a given time, called the transcriptome. These include both protein-coding and non-coding RNAs, as well as transcripts with splice and alternative initiator and terminator sites (K-H Liang, 2013). Understanding them is important for studying functional elements of the genome (Wang et al., 2009).

RNA sequencing has become a common and ubiquitous tool for analyzing quantitative changes in gene expression between experimental groups (differential gene expression or DGE) (Young et al., 2012) or in longitudinal sampling of tissues and microorganisms (Hubbard et al., 2013). The first form of RNA sequencing was used in 1977 when Fredrick Sanger developed the chain-end method, which is the first-generation sequencing platform (Adams et al., 1995, 1991). An automated version of this method was developed in 1986 (Heather and Chain, 2016). This method was based on the chemical cleavage or degradation of molecules. Innovations in microfluidics and nanotechnology have brought the era of next-generation sequencing platforms. New generation

sequencing (NGS) or second generationsequencing platforms offer a key advantage over classical Sanger sequencing, as they do not require bacterial cloning and electrophoretic separation. In 2005, Roche's model 454 was the first, a synthesis-based bioluminescence method (Metzker, 2010; van Dijk et al., 2018). In this reaction, dNTPs are cyclically added, and pyrophosphate released upon incorporation is detected. Roche 454 had a significant advantage in long reads (~1 kb) but a disadvantage in low coverage. Other technology-based developments have been initiated, resulting in the release of the Ion PGM platform in 2010 (Liu et al., 2012), which works on the basis that when a polymerase incorporates a nucleotide into DNA, a proton is released, causing a pH change that can be detected. Illumina was launched in 2007 and is currently the most widely used platform in genomics (Turnbull et al., 2018; Weimer, 2017). The sequencing process uses fluorescently labeled dNTPs, whose fluorescence is detected by a camera upon incorporation (van Dijk et al., 2014). Second-generation technologies are not suitable for the detection of long RNA molecules and their isoforms (Byrne et al., 2019).

Currently, two large sequencing platforms, PacBio and ONT MinION, are available. The former is based on nanosensor technology and has the advantage that a molecule can be sequenced several times, as the template is a circular molecule and the system does not require amplification. The system does not require amplificationof the starting material (Coupland et al., 2012). Their ONT MinION paltfrom sequencing approach is based on a completely new approach using protein nanopores embedded in a synthetic membrane (Lu et al., 2016). The sensors detect changes in ionic current corresponding to the characteristics of each nucleotide passing through. This information provides the signal used for base mapping.

Direct RNA sequencing developed by nanopore technology, which directly sequences RNA, meaning that neither reverse transcription nor PCR amplification is required for library generation, provides an efficient method to avoid major problems during transcriptome sequencing, these problems are false priming and temple shifting. These problems can lead to the detection of false

intron and transcription end sites (Cocquet et al., 2006; Garalde et al., 2018).

Viral transcripts can overlap with each other (Boldogköi, 2012; Boldogkői et al., 2019b). these overlaps can be convergent, divergent and parallel. These overlaps may play a potential role in gene regulation, meaning that they may regulate and synchronize the kinetics of viral genes through the physical interaction of transcriptional machinery. This hypothesis is called the Transcription Interference Network (TIN). Another interesting phenomenon may be generated by replication-associated RNAs transcribed near replication origins. Several such molecules have previously been detected in herpesviruses using Long-Read Sequencing (LRS) (Boldogkői et al., 2019a; Torma et al., 2023). These RNAs are thought to regulate the initiation of replication and the orientation of the replication fork. DNA replication and transcription processes are likely to generate genome-wide interference, where these two processes tightly regulate each other at the genomic level.

Epstein–Barr virus

The Epstein-Barr virus (EBV, human gammaherpesvirus 4) is a member of the subfamily Gammaherpesvirinae within the family Herpesviridae (Davison et al., 2009). EBV is predominantly transmitted by saliva and is widespread in human populations (Rickinson A, 2007). 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 (Shannon-Lowe and Rickinson, 2019; Young et al., 2016). EBV is classified as a group 1 carcinogen in humans (de Martel et al., 2020). Moreover, EBV reactivation is considered a major cause of long COVID symptoms (Gold et al., 2021). Only a limited number of viral genes from the circular, episomal, chromatinized EBV genome are expressed in latently infected cells and can be classified into three gene expression programs (types I, II and III) (Price and Luftig, 2015). In latently infected cells, viruses are replicated once per cell cycle by the host DNA synthesis machinery, with viral episomes binding to the nuclear matrix at oriP, the

latent origin of replication (Hammerschmidt and Sugden, 2013). Induction of EBV lytic replication results in a change in the limited latent expression pattern of EBV genes by sequential transcription of immediate early (IE), early (E) and late (L) EBV genes. The immediate early and early genes will be responsible for coding proteins involved in replication, while the late genes will code for the structural proteins of the virus. The EBV genom is approximately 170 kb long and actively transcribed during the lytic cycle and can encode more than 100 different gene products, of which 69 EBV-encoded proteins have recently been identified by proteomic analysis (Arvey et al., 2012; Dresang et al., 2011; Ersing et al., 2017; Yuan et al., 2006). It has been shown that transcription of the lytic cycle is bidirectional and that several newly identified transcribed regions do not encode proteins (Cao et al., 2015; Majerciak et al., 2018; O'Grady et al., 2016, 2014).

AcMNPV, a baculovirus

The Autographa californica multiple nucleopolyhedrovirus (AcMNPV) is an insect virus of the family Baculoviridae (Blissard and Rohrmann, 1990) The

developed recombinant SARS-CoV-2 antibody nanoparticle vaccine is based on this virus (Tian et al., 2021). During infection, two different forms of the virion are produced: occlusion-derived viruses, surrounded by an envelope containing viral proteins that ensure survival even in harsh environments such as the midgut of insects, and budding viruses, which have an envelope and some proteins facilitating their systemic spread in the near-neutral environment of insect tissue (Volkman et al., 1976). The 134 kbp long, double-stranded circular viral DNA contains 150 closely spaced open reading frames (ORFs) (Ayres et al., 1994). As shown by our group and others, the proximity of ORFs causes overlap between several transcripts of AcMNPV (Chen et al., 2013; Moldován et al., 2018). The AcMNPV genes are expressed in three phases: early (E), late (L), and very late (VL) (George F Rohrmann., 2008). Early transcription (0-6 hours post-infection, (p.i.)) produces transcriptional activators (Guarino and Summers, 1986) and the molecular machinery for DNA replication (Kool et al., 1994). E genes are transcribed by host RNA polymerase II. The L phase starts at the beginning of genome replication (6-18

h p.i.). The viral RNA polymerase (RNP) transcribes the L and VL genes, which recognize the consensus late initiator sequence (TAAG) on DNA and start synthesizing RNAs from the second nucleotide of the motif (Chen et al., 2013; Garrity et al., 1997). VL gene expression (18-72 h p.i.) is characterized by the synthesis of occlusion body proteins. The majority of AcMNPV transcripts contain a canonical polyadenylation signal (PAS) upstream of their transcription end site (TES).

Aims

For each virus, we aimed at six things: Identification of the 5' end of mRNAs. Determination of the 3' end of the mRNAs. Detection of promoter elements (TATA box, CAAT box, GC box) and polyadenylation signals. Linking annotated TSS and TES positions to transcripts. Categorization and abundance determination of annotated transcript isoforms, polycistronic RNAs, ncRNAs, antisense RNAs and 5' truncated RNAs. Detection of transcriptional overlaps

Materials and methods

Viruses were propagated in appropriate cell lines grown in conditions recommended by the cell line vendor. The infection was stopped in consecutive time points, and total RNA was extracted from the cell lysates. For EBV total RNA samples were split in two. Polyadenylated RNAs were isolated from half of the total RNA. Ribodepletion, was carried out to remove ribosomal RNA from the other half of total RNAs. In case of AcMNPV polyadenylated RNAs and Cap-selected RNAs were isolated Futhermore, RNA bisulfite conversion was carried out. MinION amplified full-length cDNA libraries were prepared for all viruses, while direct RNA libraries was prepared for the AcMNPV and a non-amplified full-length cDNA library was prepared for EBV, followed by sequencing on PacBio and MinION sequencers.

Our sequencing data was base called with Minimap2 and pre-processed, followed by transcript isoform annotation using the LoRTIA software suite. The IGV and Genious software tools were employed for visualization of the reads and annotated transcripts.

Results

In case of AcMNPV, we identified 311 transcription start sites (TSS) and 261 transcription end sites (TES), from which we annotated a total of 875 transcripts. Among these, 330 5'-UTR isoforms, as well as 340 3'-UTR variants, were determined, 76.35% of which contained canonical PAS upstream of their 5′-end. Additionally, we identified 41 5'-truncated transcripts, which are RNAs containing in-frame ORFs within genes. Multicistronic transcripts that overlap multiple genes were also identified, with 45 being polycistronic and 54 being complex transcripts. We identified 101 novel non-coding transcript isoforms that did not contain previously annotated ORFs, two-thirds of which were longer than 200 nts, representing long non-coding RNAs. The AcMNPV genome contains 37 convergent gene pairs. Our LRS analysis revealed that all convergent gene pairs exhibited transcriptional readthroughs. Among these, only three pairs exclusively overlapped in their 3'-UTRs, while the remaining pairs exhibited overlaps in their ORFs. Out of the 34 gene pairs, 32 showed divergent transcriptional overlaps, and 84 demonstrated parallel overlaps in 87 gene

pairs. Homologous repeat (hr) regions are located at multiple genomic positions in AcMNPV, believed to contain replication origins (Oris). In total, 55 transcript species were identified in hr regions. RNA modification detection was performed on the direct RNA sequencing data of AcMNPV using the Tombo software suite. After filtering the potential false-positive detections 325 putative 5 m C sites mapped to the viral transcriptome in 12 viral genes. We detected a potential methylation consensus sequence: UUACCG (the modified base underlined). With bisulfite sequencing, 234 of the 325 methylated positions (identified by Tombo analysis) were confirmed In total, 7897 putative methylation positions were identified in 99 gene transcripts. 31 potential cytosine positions were detected in the 3′-UTR of the ac-Orf-12 transcript, all of which were untranslated and therefore methylated. Overall, 88% of the potential methylation positions tested were located in coding regions and 21% in UTRs. Reads of ORF19-L showed a high frequency of A to I (read A to G by sequencing) substitution, which was not present in the overlapping

reads. We found that for ORF19, 50% of all substitutions were A to G.

For EBV, 322 TSS (145 novel) and 57 TES (12 novel) were identified, from which a total of 351 transcripts were annotated from them 241 where novel. This included 47 long and 57 short 5'-UTR isoforms, along with 7 3' UTR isoforms. 72 5'-truncated transcripts were identified. Using the LoRTIA toolkit, 205 introns were detected from which 42 novel spliced transcripts were detected. We identified forty-seven multigenic transcripts, 27 of which were novel, and 4 novel complex transcripts out of a total of 6. A total of 21 non-coding transcripts were detected, 2 short ncRNAs and 19 lncRNAs. All three forms of transcriptional overlap were detected in EBV RNAs. The EBV genome has two lytic (Ori-Lyt) and one latent (OriP) replication origins. We detected nine novel isoforms of Ori-associated RNAs.

Discusion

In my thesis I describe the analysis of the transcriptome of AcMNPV and EBV using sequencing platforms that provide long reads. Using these approaches ,

several new, previously unknown transcript categories have been identified. As a result of our investigations, we have identified numerous new transcription start sites and transcription end points. Additionally, we have discovered new 5' and 3'-UTR isoforms. The Long-Read Sequencing (LRS) revealed a broad spectrum of 5'-truncated hypothetical RNAs. Several multigenic transcripts were detected, that overlap multiple genes in each virus. It is hypothesized that transcriptional readthrough in tandem genes (and also on convergent genes) plays a role in a transcription interference-based mechanism (Boldogköi, 2012). In this study, we detected novel promoters, Inr sequences, and poly(A) sites. Another significant result of these platforms was that a new class of RNAs was identified with them, which include transcripts transcribing replication origins and transcripts near replication origins (Boldogkői et al., 2019a). In conclusion, multiplatform approaches are essential in transcriptomic studies because different platforms have distinct advantages and limitations, representing independent techniques vital for validating results obtained by a particular method.

Reference

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., Venter, J.C., 1991. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. Science (80-.). 252, 1651–1656. https://doi.org/10.1126/science.2047873

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377, 3–174.

Arvey, A., Tempera, I., Tsai, K., Chen, H.S., Tikhmyanova, N., Klichinsky, M., Leslie, C., Lieberman, P.M., 2012. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. Cell Host Microbe 12, 233–245. https://doi.org/10.1016/j.chom.2012.06.008

Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The Complete DNA Sequence of Autographa californica Nuclear Polyhedrosis Virus. Virology 202, 586–605. https://doi.org/10.1006/viro.1994.1380

Blissard, G.W., Rohrmann, G.F., 1990. Baculovirus Diversity and Molecular Biology. Annu. Rev. Entomol. 35, 127–155. https://doi.org/10.1146/annurev.en.35.010190.001015

Boldogköi, Z., 2012. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. Front. Genet. 3, 1–17. https://doi.org/10.3389/fgene.2012.00122

Boldogkői, Z., Balázs, Z., Moldován, N., Prazsák, I., Tombácz, D., 2019a. Novel classes of replication-associated transcripts discovered in viruses. RNA Biol. 16, 166–175.
https://doi.org/10.1080/15476286.2018.1564468

Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M., Tombácz, D., 2019b. Long-Read Sequencing – A

Powerful Tool in Viral Transcriptome Research. Trends Microbiol. 27, 578–592. https://doi.org/10.1016/j.tim.2019.01.010

Byrne, A., Cole, C., Volden, R., Vollmers, C., 2019. Realizing the potential of full-length transcriptome sequencing. Philos. Trans. R. Soc. B Biol. Sci. 374, 20190097. https://doi.org/10.1098/rstb.2019.0097

Cao, S., Moss, W., O'Grady, T., Concha, M., Strong, M.J., Wang, X., Yu, Y., Baddoo, M., Zhang, K., Fewell, C., Lin, Z., Dong, Y., Flemington, E.K., 2015. 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. https://doi.org/10.1128/jvi.00608-15

Chen, Y.-R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J.Z., Zhang, S., Blissard, G.W., 2013. The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in Trichoplusia ni Cells. J.

Virol. 87, 6391–6405. https://doi.org/10.1128/JVI.00194-13

Cocquet, J., Chong, A., Zhang, G., Veitia, R.A., 2006. Reverse transcriptase template switching and false alternative transcripts. Genomics 88, 127–131. https://doi.org/10.1016/j.ygeno.2005.12.013

Coupland, P., Chandra, T., Quail, M., Reik, W., Swerdlow, H., 2012. Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. Biotechniques 53, 365–372. https://doi.org/10.2144/000113962

Davison, A.J., Eberle, R., Ehlers, B., Hayward, G.S., McGeoch, D.J., Minson, A.C., Pellett, P.E., Roizman, B., Studdert, M.J., Thiry, E., 2009. The order Herpesvirales. Arch. Virol. 154, 171–177. https://doi.org/10.1007/s00705-008-0278-4

de Martel, C., Georges, D., Bray, F., Ferlay, J., Clifford, G.M., 2020. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis.

Lancet Glob. Heal. 8, e180–e190. https://doi.org/10.1016/S2214-109X(19)30488-7

Dresang, L.R., Teuton, J.R., Feng, H., Jacobs, J.M., Camp, D.G., Purvine, S.O., Gritsenko, M.A., Li, Z., Smith, R.D., Sugden, B., Moore, P.S., Chang, Y., 2011. Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes. BMC Genomics 12, 625. https://doi.org/10.1186/1471-2164-12-625

Ersing, I., Nobre, L., Wang, L.W., Soday, L., Ma, Y., Paulo, J.A., Narita, Y., Ashbaugh, C.W., Jiang, C., Grayson, N.E., Kieff, E., Gygi, S.P., Weekes, M.P., Gewurz, B.E., 2017. A Temporal Proteomic Map of Epstein-Barr Virus Lytic Replication in B Cells. Cell Rep. 19, 1479–1493. https://doi.org/10.1016/j.celrep.2017.04.062

Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E.J., Jayasinghe, L., Wright, C., Blasco, J., Young, S., Brocklebank, D., Juul,

S., Clarke, J., Heron, A.J., Turner, D.J., 2018. Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods 15, 201–206. https://doi.org/10.1038/nmeth.4577

Garrity, D.B., Chang, M.-J., Blissard, G.W., 1997. Late Promoter Selection in the Baculovirusgp64 Envelope Fusion ProteinGene. Virology 231, 167–181. https://doi.org/10.1006/viro.1997.8540

George F Rohrmann., 2008. MOLECULAR BIOLOGY Molecular Biology General. ethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information, Oregon.

Gold, J.E., Okyay, R.A., Licht, W.E., Hurley, D.J., 2021. Investigation of long covid prevalence and its relationship to epstein-barr virus reactivation. Pathogens 10, 1–15. https://doi.org/10.3390/pathogens10060763

Guarino, L.A., Summers, M.D., 1986. Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. J. Virol. 57, 563–571. https://doi.org/10.1128/jvi.57.2.563-571.1986

Hammerschmidt, W., Sugden, B., 2013. Replication of Epstein-Barr Viral DNA. Cold Spring Harb. Perspect. Biol. 5, a013029–a013029. https://doi.org/10.1101/cshperspect.a013029

Heather, J.M., Chain, B., 2016. The sequence of sequencers: The history of sequencing DNA. Genomics 107, 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hubbard, K.S., Gut, I.M., Lyman, M.E., McNutt, P.M., 2013. Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. F1000Research 2, 35. https://doi.org/10.12688/f1000research.2-35.v1

K-H Liang, 2013. Bioinformatics for Biomedical Science and Clinical Applications, 1 st. ed. Woodhead Publishing.

Kool, M., Ahrens, C.H., Goldbach, R.W., Rohrmann, G.F., Vlak, J.M., 1994. Identification of genes involved in DNA replication of the Autographa californica baculovirus. Proc. Natl. Acad. Sci. 91, 11212–11216. https://doi.org/10.1073/pnas.91.23.11212

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of Next-Generation Sequencing Systems. J. Biomed. Biotechnol. 2012, 1–11. https://doi.org/10.1155/2012/251364

Lu, H., Giordano, F., Ning, Z., 2016. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics. Proteomics Bioinformatics 14, 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Majerciak, V., Yang, W., Zheng, J., Zhu, J., Zheng, Z.-M., 2018. A Genome-Wide Epstein-Barr Virus Polyadenylation Map and Its Antisense RNA to EBNA. J. Virol. 93. https://doi.org/10.1128/jvi.01593-18

Metzker, M.L., 2010. Sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46. https://doi.org/10.1038/nrg2626

Moldován, N., Tombácz, D., Szucs, A., Csabai, Z., Balázs, Z., Kis, E., Molnár, J., Boldogkoi, Z., Szűcs, A., Csabai, Z., Balázs, Z., Kis, E., Molnár, J., Boldogkői, Z., 2018. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a

Baculovirus. Sci. Rep. 8, 8604. https://doi.org/10.1038/s41598-018-26955-8

O'Grady, T., Cao, S., Strong, M.J., Concha, M., Wang, X., Splinter Bondurant, S., Adams, M., Baddoo, M., Srivastav, S.K., Lin, Z., Fewell, C., Yin, Q., Flemington, E.K., 2014. Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. J. Virol. 88, 1604–16. https://doi.org/10.1128/JVI.02989-13

O'Grady, T., Wang, X., Höner zu Bentrup, K., Baddoo, M., Concha, M., Flemington, E.K., Höner zu Bentrup, K., Baddoo, M., Concha, M., Flemington, E.K., 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Res. 44, e145–e145. https://doi.org/10.1093/nar/gkw629

Price, A.M., Luftig, M.A., 2015. To Be or Not IIb: A Multi-Step Process for Epstein-Barr Virus Latency Establishment and Consequences for B Cell Tumorigenesis. PLOS Pathog. 11, e1004656. https://doi.org/10.1371/journal.ppat.1004656

Rickinson A, K., 2007. Epstein-Barr Virus, in: D., K., P., H. (Eds.), Fields Virology. Lippincott, Williams and Wilkins, Philadelphia, pp. 2655–2700.

Shannon-Lowe, C., Rickinson, A., 2019. The Global Landscape of EBV-Associated Tumors. Front. Oncol. https://doi.org/10.3389/fonc.2019.00713

Tian, J.-H., Patel, N., Haupt, R., Zhou, H., Weston, S., Hammond, H., Logue, J., Portnoff, A.D., Norton, J., Guebre-Xabier, M., Zhou, B., Jacobson, K., Maciejewski, S., Khatoon, R., Wisniewska, M., Moffitt, W., Kluepfel-Stahl, S., Ekechukwu, B., Papin, J., Boddapati, S., Jason Wong, C., Piedra, P.A., Frieman, M.B., Massare, M.J., Fries, L., Bengtsson, K.L., Stertman, L., Ellingsworth, L., Glenn, G., Smith, G., 2021. SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 immunogenicity in baboons and protection in mice. Nat. Commun. 12, 372. https://doi.org/10.1038/s41467-020-20653-8

Torma, G., Tombácz, D., Csabai, Z., Almsarrhad, I.A.A., Nagy, G.Á., Kakuk, B., Gulyás, G., Spires, L.M., Gupta, I., Fülöp, Á., Dörmő, Á., Prazsák, I., Mizik, M.,

Dani, V.É., Csányi, V., Harangozó, Á., Zádori, Z., Toth, Z., Boldogkői, Z., 2023. Identification of herpesvirus transcripts from genomic regions around the replication origins. Sci. Rep. 13, 16395. https://doi.org/10.1038/s41598-023-43344-y

Turnbull, C., Scott, R.H., Thomas, E., Jones, L., Murugaesu, N., Pretty, F.B., Halai, D., Baple, E., Craig, C., Hamblin, A., Henderson, S., Patch, C., O'Neill, A., Devereau, A., Smith, K., Martin, A.R., Sosinsky, A., McDonagh, E.M., Sultana, R., Mueller, M., Smedley, D., Toms, A., Dinh, L., Fowler, T., Bale, M., Hubbard, T., Rendon, A., Hill, S., Caulfield, M.J., 2018. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ k1687. https://doi.org/10.1136/bmj.k1687

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends Genet. 30, 418–426. https://doi.org/10.1016/j.tig.2014.07.001

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., Thermes, C., 2018. The Third Revolution in Sequencing

Technology. Trends Genet. 34, 666–681. https://doi.org/10.1016/j.tig.2018.05.008

Volkman, L.E., Summers, M.D., Hsieh, C.H., 1976. Occluded and nonoccluded nuclear polyhedrosis virus grown in Trichoplusia ni: comparative neutralization comparative infectivity, and in vitro growth studies. J. Virol. 19, 820–832. https://doi.org/10.1128/jvi.19.3.820-832.1976

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. https://doi.org/10.1038/nrg2484

Weimer, B.C., 2017. 100K Pathogen Genome Project. Genome Announc. 5. https://doi.org/10.1128/genomeA.00594-17

Young, L.S., Yap, L.F., Murray, P.G., 2016. Epstein-Barr virus: More than 50 years old and still providing surprises. Nat. Rev. Cancer. https://doi.org/10.1038/nrc.2016.92

Young, M.D., McCarthy, D.J., Wakefield, M.J., Smyth, G.K., Oshlack, A., Robinson, M.D., 2012.

Differential Expression for RNA Sequencing (RNA-Seq) Data: Mapping, Summarization, Statistical Analysis, and Experimental Design, in: Bioinformatics for High Throughput Sequencing. Springer New York, New York, NY, pp. 169–190. https://doi.org/10.1007/978-1-4614-0782-9_10

Yuan, J., Cahir-McFarland, E., Zhao, B., Kieff, E., 2006. Virus and Cell RNAs Expressed during Epstein-Barr Virus Replication. J. Virol. 80, 2548–2565. https://doi.org/10.1128/jvi.80.5.2548-2565.2006