

**Emberi és rovar DNS-vírus transzkriptomikai  
elemzése integrált szekvenálási megközelítéssel**

**Ph.D. tézisfüzet**

**Fülöp Ádám**



**Interdiszciplináris Orvostudományok Doktori Iskola**

**Általános Orvostudományi Kar**

**Szegedi Tudományegyetem**

**Témavezetők: Prof. Dr. Boldogkői Zsolt, Ph. D, DSc**

**Szeged**

**- 2024 -**

## **A tézishez közvetlenül kapcsolódó publikációk listája**

**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**

## **Bevezető**

Az adott időpontban kifejezett összes RNS-molekula, az úgynevezett transzkriptom. Ezek közé tartoznak mind a fehérjekódoló, mind a nem kódoló RNS-ek, valamint a splice- és alternatív iniciátor- és terminátorhelyekkel rendelkező transzkriptek (K-H Liang, 2013). Ezek megértése fontos a genom funkcionális elemeinek tanulmányozásához (Wang et al., 2009).

Az RNS-szekvenálás elterjedt és mindenütt jelen lévő eszközzé vált a génexpresszióban a kísérleti csoportok közötti mennyiségi változások (differenciális génexpresszió vagy DGE) (Young et al., 2012) vagy a szövetek és mikroorganizmusok longitudinális mintavételezése során (Hubbard et al., 2013). Az RNS-szekvenálás első formáját 1977-ben alkalmazták, amikor Fredrick Sanger kifejlesztette a láncvéges módszert, amely az első generációs szekvenálási platform (Adams et al., 1995, 1991). A módszer automatizált változatát 1986-ban fejlesztették ki (Heather and Chain, 2016). Ez a módszer a molekulák kémiai hasításán vagy lebontásán alapult. A mikrofluidikában és a nanotechnológiában bekövetkezett innovációk elhozták az újgenerációs szekvenálási

platformok korszakát. Az újgenerációs szekvenálási (NGS) vagy második generációs szekvenálási platformok kulcsfontosságú előnyt kínálnak a klasszikus Sanger-szekvenálással szemben, mivel nem igényelnek bakteriális klónozást és elektroforetikus elválasztást. 2005-ben a Roche 454-es modellje volt az első, egy szintézis alapú biolumineszcens módszer (Metzker, 2010; van Dijk et al., 2018). Ebben a reakcióban a dNTP-eket ciklikusan adagolják, és a beépüléskor felszabaduló pirofoszfátot detektálják. A Roche 454 jelentős előnyt jelentett a hosszú leolvasások (~1 kb) esetében, de hátrányt az alacsony lefedettségben. További technológiai alapú fejlesztések indultak, amelyek eredményeként 2010-ben megjelent az Ion PGM platform (Liu et al., 2012), amely azon az alapon működik, hogy amikor a polimeráz beépít egy nukleotidot a DNS-be, proton szabadul fel, ami pH-változást okoz, ami detektálható. Az Illumina 2007-ben indult el, és jelenleg a genomikában a legszélesebb körben használt platform (Turnbull et al., 2018; Weimer, 2017). A szekvenálás során fluoreszcensen jelölt dNTP-eket használnak, amelyek beépüléskor felszabaduló fluoreszcenciát egy kamera érzékeli (van Dijk et al., 2014).

A második generációs technológiák nem alkalmasak hosszú RNS-molekulák és izoformáik kimutatására (Byrne et al., 2019).. Jelenleg két nagy szekvenáló platform, a PacBio és az ONT MinION áll rendelkezésre. Az előbbi nanoszenzoros technológián alapul, és előnye, hogy egy molekula többször is szekvenálható, mivel a templát körkörös molekula, és a rendszer nem igényel amplifikációt. A rendszer nem igényli a kiindulási anyag amplifikációját (Coupland et al., 2012). ONT MinION platform szekvenálási módszerük egy teljesen új megközelítésen alapul, amely szintetikus membránba ágyazott fehérje nanopórusokat használ (Lu et al., 2016). A szenzorok az egyes áthaladó nukleotidok jellemzőinek megfelelő ionáram-változásokat érzékelik. Ez az információ adja a bázistérképezéshez használt jelet.

A nanopore technology kifejlesztett közvetlen RNS-szekvenálást, amely közvetlenül szekvenálja az RNS-t, ami azt jelenti, hogy a könyvtár létrehozásához nincs szükség sem reverz transzkripcióra, sem PCR-amplifikációra, hatékony módszer a transzkriptom szekvenálás során felmerülő főbb problémák - a hamis

priming és a templáteltolódás - elkerülésére. Ezek a problémák hamis intron- és transzkripciós véghelyek detektálásához vezethetnek (Cocquet et al., 2006; Garalde et al., 2018).

A vírusos transzkriptumok átfedhetnek egymással (Boldogkői, 2012; Boldogkői et al., 2019b). ezek az átfedések lehetnek konvergensek, divergensek és párhuzamosak. Ezek az átfedések potenciális szerepet játszhatnak a génszabályozásban, vagyis a transzkripciós gépezet fizikai kölcsönhatásán keresztül szabályozhatják és szinkronizálhatják a vírusgének kinetikáját. Ezt a hipotézist nevezzük transzkripciós interferenciahálózatnak (TIN). Egy másik érdekes jelenséget a replikációs origók közelében átíródó replikáció-asszociált RNS-ek hozhatnak létre. Korábban több ilyen molekulát is kimutattak herpeszvírusokban hosszú leolvasású szekvenálással (LRS) (Boldogkői et al., 2019a; Torma et al., 2023). Ezek az RNS-ek feltehetően a replikáció megindulását és a replikációs villa orientációját szabályozzák. A DNS-replikációs és transzkripciós folyamatok valószínűleg genom szintű interferenciát

generálnak, ahol e két folyamat genomi szinten szorosan szabályozza egymást.

### **Epstein–Barr vírus**

Az Epstein-Barr-vírus (EBV, human gammaherpesvirus 4) a Herpesviridae családon belül a Gammaherpesvirinae alcsalád tagja (Davison et al., 2009). Az EBV túlnyomórészt nyállal terjed, és széles körben elterjedt az emberi populációkban (Shannon-Lowe and Rickinson, 2019; Young et al., 2016). Az EBV szerepet játszik a Burkitt-limfóma és más limfómák patogenezisében, valamint részt vesz a nazofaringeális karcinóma és a gyomorkarcinómák egy alcsoportjának kialakulásában is (Shannon-Lowe és Rickinson, 2019; Young és mtsai., 2016). Az EBV-t az emberben az 1. csoportba tartozó rákkeltő anyagnak minősítik (de Martel et al., 2020). Ezenkívül az EBV reaktiválódását tartják a hosszú COVID-tünetek egyik fő okának (Gold et al., 2021). A cirkuláris, epizomális, kromatinizált EBV genomból csak korlátozott számú vírusgén fejeződik ki a látensen fertőzött sejtekben, és három génexpressziós programba sorolható (I., II. és III. típus) (Price and Luftig,

2015). A látens fertőzött sejtekben a vírusok sejtciklusonként egyszer replikálódnak a gazdaszervezet DNS-szintézis gépezete által, a vírus episzómák a nukleáris mátrixhoz kötődnek az oriP-nél, a látens replikációs origónál (Hammerschmidt and Sugden, 2013). Az EBV lítikus replikációjának indukciója az EBV-gének korlátozott látens expressziós mintázatának megváltozását eredményezi az azonnali korai (IE), korai (E) és késői (L) EBV-gének szekvenciális transzkripciója révén. Az azonnali korai és korai gének a replikációban részt vevő fehérjék kódolásáért lesznek felelősek, míg a késői gének a vírus strukturális fehérjéit kódolják. Az EBV genomja körülbelül 170 kb hosszú, a lítikus ciklus során aktívan átíródik, és több mint 100 különböző génterméket kódolhat, amelyek közül a közelmúltban 69 EBV által kódolt fehérjét azonosítottak proteomikai elemzéssel (Arvey et al., 2012; Dresang et al., 2011; Ersing et al., 2017; Yuan et al., 2006). Kimutatták, hogy a lítikus ciklus transzkripciója kétirányú, és számos újonnan azonosított átírt régió nem kódol fehérjéket (Cao et al., 2015; Majerciak et al., 2018; O'Grady et al., 2016, 2014).



## **AcMNPV, a baculovirus**

Az *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) a Baculoviridae családba tartozó rovarvírus (Blissard and Rohrmann, 1990). A kifejlesztett rekombináns SARS-CoV-2 antitest nanorészecskés vakcina ezen a víruson alapul (Tian et al., 2021). A fertőzés során a virion két különböző formája keletkezik: az okklúziós vírusok, amelyeket egy vírusfehérjéket tartalmazó burok vesz körül, amely még a zord környezetben, például a rovarok középbelrendszerében is biztosítja a túlélést, és a budding vírusok, amelyeknek a burok és néhány fehérje megkönnyíti a szisztémás terjedésüket a rovarszövetek közel semleges környezetében (Volkman et al., 1976). A 134 kbp hosszú, kettős szálú, cirkuláris vírus-DNS 150 szorosan egymás mellett elhelyezkedő nyitott olvasókeretet (ORF) tartalmaz (Ayres et al., 1994). Amint azt csoportunk és mások is kimutatták, az ORF-ek közelsége átfedést okoz az AcMNPV több transzkriptuma között (Chen et al., 2013; Moldován et al., 2018). Az AcMNPV gének három fázisban fejeződnek ki: korai (E),

késői (L) és nagyon késői (VL) fázisban (George F Rohrmann., 2008). A korai transzkripció (0-6 órával a fertőzés után, (p.i.)) transzkripciós aktivátorokat (Guarino and Summers, 1986) és a DNS-replikáció molekuláris gépezetét termeli (Kool et al., 1994). Az E géneket a gazdaszervezet RNS-polimeráz II írja át. Az L fázis a genom replikáció kezdetén kezdődik (6-18 h p.i.). A vírus RNS-polimeráz (RNP) átírja az L és VL géneket, amelyek felismerik a DNS-en lévő konszenzusos késői iniciátor szekvenciát (TAAG), és a motívum második nukleotidjától kezdik az RNS-ek szintézisét (Chen et al., 2013; Garrity et al., 1997). A VL gének expresszióját (18-72 h p.i.) az okklúziós test fehérjék szintézise jellemzi. Az AcMNPV transzkriptumok többsége tartalmaz egy kanonikus poliadenilációs jelet (PAS) a transzkripciós véghelyük (TES) előtt.

### **Célkitűzések**

Minden egyes vírus esetében hat dolgot céloztunk meg: Az mRNS-ek 5' végének azonosítása. Az mRNS-ek 3' végének meghatározása. Promóter elemek (TATA box, CAAT box, GC box) és poliadenilációs jelek kimutatása.

Annotált TSS és TES pozíciók összekapcsolása a transzkriptekkel. Annotált transzkript izoformák, policisztronikus RNS-ek, ncRNS-ek, antiszenz RNS-ek és 5' csonka RNS-ek kategorizálása és abundancia meghatározása. A transzkripció átfedések kimutatása

### **Anyagok és módszerek**

A vírusokat megfelelő sejtvonalakban szaporítottuk, amelyeket a sejtvonal szállítója által ajánlott körülmények között tenyésztettünk. A fertőzést egymást követő időpontokban leállítottuk, és a sejtlyúzatumokból teljes RNS-t kinyerettünk. Az EBV esetében a teljes RNS-mintákat két részre osztottuk. A teljes RNS feléből poliadenilált RNS-t izoláltunk. Ribodepléciót végeztünk a riboszómális RNS eltávolítására a teljes RNS-ek másik feléből. Az AcMNPV esetében poliadenilált RNS-eket és Cap-szelektált RNS-eket izoláltunk. Minden vírus esetében MinION-amplifikált teljes hosszúságú cDNS-könyvtárat készítettünk, míg az AcMNPV esetében direkt RNS-könyvtárat, az EBV esetében pedig nem amplifikált teljes hosszúságú cDNS-könyvtárat, majd szekvenálást végeztünk PacBio és MinION szekvenátorokon.

A szekvenálási adatainkat a Minimap2 segítségével báziskódoltuk és előfeldolgoztuk, majd a LoRTIA szoftvercsomag segítségével transzkript izoformák annotációját végeztük el. Az IGV és a Genious szoftvereszközöket használtuk az leolvasások és az annotált transzkriptek vizualizálásához.

## **Eredmények**

Az AcMNPV esetében 311 transzkripció starthelyet (TSS) és 261 transzkripció véghelyet (TES) azonosítottunk, amelyekből összesen 875 transzkriptumot annotáltunk. Ezek közül 330 5'-UTR izoformát, valamint 340 3'-UTR variánst határoztunk meg, amelyek 76,35%-a tartalmazott 5'-végük előtt kanonikus PAS-t. Ezen kívül 41 5'-csonkolt transzkriptet azonosítottunk, amelyek olyan RNS-ek, amelyek a géneken belül in-frame ORF-eket tartalmaznak. Azonosítottuk a több gént átfedő multicisztronikus transzkriptumokat is, amelyek közül 45 policisztronikus és 54 komplex transzkriptum volt. Azonosítottunk 101 új, nem kódoló transzkript izoformát, melyek nem tartalmaztak korábban annotált ORF-eket, kétharmaduk 200 nts-nál hosszabb volt, és a hosszú, nem

kódoló RNS-eket képviselte. Az AcMNPV genom 37 konvergens génpárt tartalmaz. LRS-elemzésünk kimutatta, hogy minden konvergens génpár transzkripció átírást mutatott. Ezek közül csak három pár fedte át kizárólag a 3'-UTR-jeiket, míg a többi pár az ORF-jeikben mutatott átfedéseket. A 34 génpárból 32 mutatott divergens transzkripció átfedést, és 84 mutatott párhuzamos átfedést 87 génpárban. A homológ ismétlődő (hr) régiók több genomiális pozícióban helyezkednek el az AcMNPV-ben, amelyekről feltételezhető, hogy replikációs origókat (Ori-k) tartalmaznak. Összesen 55 transzkriptfajt azonosítottunk a hr régiókban. Az AcMNPV közvetlen RNS-szekvenálási adatain RNS-módosítások kimutatását végeztük a Tombo szoftvercsomag segítségével. A potenciális álpozitív detektálások kiszűrése után 325 feltételezett 5 m C-helyet térképeztünk fel a vírus transzkriptomjában 12 vírusgénben. Egy potenciális metilációs konszenzus szekvenciát detektáltunk: UUACCG (a módosított bázis aláhúzva). Biszulfid-szekvenálással a 325 metilált pozícióból 234-et (a Tombo-elemzéssel azonosított) megerősítettünk. Összesen 7897 feltételezett metilációs pozíciót azonosítottunk 99 gén

transzkriptumban. Az ac-Orf-12 transzkriptum 3'-UTR-ében 31 potenciális citozin-pozíciót detektáltunk, amelyek egyike sem transzlálódott és ezért metilált volt. Összességében a vizsgált potenciális metilációs pozíciók 88%-a a kódoló régiókban, 21%-a pedig az UTR-ben helyezkedett el. Az ORF19-L leolvasásai nagy gyakorisággal mutattak A-tól I-re (szekvenálással olvasva A-tól G-re) történő szubsztitúciót, ami az átfedő leolvasásokban nem volt jelen. Megállapítottuk, hogy az ORF19 esetében az összes szubsztitúció 50%-a A-G-re változott.

Az EBV esetében 322 TSS-t (145 új) és 57 TES-t (12 új) azonosítottunk, amelyekből összesen 351 transzkriptet annotáltunk, ezek közül 241 volt új. Ez 47 hosszú és 57 rövid 5'-UTR izoformát, valamint 7 3'-UTR izoformát tartalmazott. 72 5'-csonkolt transzkriptet azonosítottunk. A LoRTIA eszközkészlet segítségével 205 intront detektáltunk, amelyekből 42 új splicing transzkriptet találtunk. Negyvenhét multigenikus transzkriptet azonosítottunk, amelyek közül 27 új volt, és 4 új komplex transzkriptet azonosítottunk az összesen 6ból.

Összesen 21 nem kódoló transzkriptet detektáltunk, 2 rövid ncRNS-t és 19 lncRNS-t. A transzkripció átfedés mindhárom formáját kimutatták az EBV RNS-ekben. Az EBV genom két lítikus (Ori-Lyt) és egy látens (OriP) replikációs origóval rendelkezik. Az Ori-asszociált RNS-ek kilenc új izoformáját mutattuk ki.

## **Diszkusszió**

Tézisemben az AcMNPV és az EBV transzkriptomjának elemzését ismertetem hosszú leolvasásokat biztosító szekvenálási platformok segítségével. Ezen megközelítések segítségével, számos új, korábban ismeretlen transzkript kategóriát azonosítottunk. Vizsgálataink eredményeként számos új transzkripcióshelyet és transzkripcióshelyet azonosítottunk. Emellett új 5' és 3'-UTR izoformákat fedeztünk fel. A hosszú leolvasású szekvenálás (LRS) az 5'-csonkolt hipotetikus RNS-ek széles spektrumát tárta fel. Számos multigenikus transzkriptumot észleltünk, amelyek minden egyes vírusban több gént átfednek. Feltételezhető, hogy a transzkripcióshely átolvasás a tandemgéneken (és a konvergens géneken is) egy transzkripcióshely interferencián alapuló mechanizmusban játszik szerepet (Boldogkői,

2012). Ebben a tanulmányban új promótereket, iniciációs szekvenciákat és poli(A)-helyeket detektáltunk. A platformok másik jelentős eredménye, hogy az RNS-ek egy új osztályát azonosítottuk velük, amelyek a replikációs origókat átíró és a replikációs origókhoz közeli transzkriptumokat tartalmazzák (Boldogkői et al., 2019a). Összefoglalva, a multiplatformos megközelítések nélkülözhetetlenek a transzkriptomikai vizsgálatokban, mivel a különböző platformok eltérő előnyökkel és korlátokkal rendelkeznek, és független technikákat képviselnek, amelyek elengedhetetlenek egy adott módszerrel kapott eredmények validálásához.



**Hivatkozásjegyzék**

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>

Boldogkői, 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.*

Viol. 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](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., Tombác, D., Szucs, A., Csabai, Z., Balázs, Z., Kis, E., Molnár, J., Boldogkői, 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>

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](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>