

**Nagy DNS vírusok analízise hosszú leolvasásokat adó RNS
szekvenálással**

Ph.D. tézisfüzet

Torma Gábor



Orvosi Biológia Intézet

Interdiszciplináris Orvostudományok Doktori Iskola

Általános Orvostudományi Kar

Szegedi Tudományegyetem

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

Dr. Tombác Dóra, Ph. D

Szeged

- 2024 -

A tézishez közvetlenül kapcsolódó publikációk listája

I. Olasz F, Tombácz D, **Torma G**, Csabai Z, Moldován N, Dörmő Á, Prazsák I, Mészáros I, Magyar T, Tamás V, Zádori Z, Boldogkői Z. Short and Long-Read

Sequencing Survey of the Dynamic Transcriptomes of African Swine Fever Virus and the Host Cells. *Front Genet.* 2020 Jul 28;11:758. doi: 10.3389/fgene.2020.00758. **IF: 4,599**

II. Torma G, Tombácz D, Csabai Z, Göbhardt D, Deim Z, Snyder M, Boldogkői Z. An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. *Pathogens.* 2021 Feb 20;10(2):242. doi: 10.3390/pathogens10020242. **IF: 4,531**

III. Torma G, Tombácz D, Csabai Z, Moldován N, Mészáros I, Zádori Z, Boldogkői Z. Combined Short and Long-Read Sequencing Reveals a Complex Transcriptomic Architecture of African Swine Fever

Virus. Viruses. 2021 Mar 30;13(4):579.
doi:10.3390/v13040579. **IF: 5,818**

Bevezető

Egy organizmus genomjában átíródó összes RNS-t transzkriptomnak nevezzük¹. Ez magában foglalja a fehérjét kódoló és nem kódoló RNS-eket (ncRNS). Megértésük fontos a genetikai szabályozások pontos megismeréséhez². A transzkriptomika fő célja ezen RNS molekulák vizsgálata az egyedfejlődés különböző stádiumaiban.

A transzkriptom vizsgálatok kezdetben Northern blottal és Real-time PCR-el valósultak meg, azonban ezek a transzkriptom egy kis részletéről képesek információt adni^{3,4}. Az áttörések kezdetét a Sanger szekvenátorok megjelenése jelentette, mivel a Humán Genom Projekt során történő felhasználásuk óriási innovációs igényt ébresztett fel nagy áteresztőképességű szekvenálási platformok kifejlesztésére^{5,6}. Ez a következő generációs szekvenálási platformok megjelenéséhez vezetett, amelyek nagy előnye a Sanger szekvenátorokkal szemben, hogy nincs szükség elektroforetikus elválasztásra és

bakteriális klónozásra⁷. Ezen szekvenátoroknak a legelső képviselője a Roche's 454 volt, melynek legnagyobb előnye a hosszú ~1kb leolvasások voltak^{7,8}. Vele párhuzamosan megjelent az Ion PGM platformja is, amely nem igényel optikai detekciót, mivel a nukleotidok beépítése során kialakuló pH változásokat detektálja⁹. Ezt követően 2007-ben jelent meg az Illumina platformja, mely jelenleg a legelterjedtebb rövid read szekvenátor^{10,11}. Használatával ugrásszerűen megnőtt az ismeretünk az egyes organizmusok szekvenciáról. A jelenlegi genom programok Illumina platformokon futnak. Előnye a nagy áteresztőképesség¹². Működése a szintéziskor beépülő fluoreszcensen jelölt nukleotidok fluoreszcenciájának detekcióján alapul⁸. Azonban ezen platform a rövid szekvencia leolvasásai miatt nem alkalmas egész hosszúságú RNS molekulák detekciójára¹³. Ezért a közelmúltban a transzkriptom analízisben megjelentek a harmadik generációs szekvenátorok, amelyek képesek egész hosszúságú RNS molekulákat végponttól végpontig detektálni. Ennek a platformnak jelenleg két képviselője van, az egyik az Oxford Nanopore által fejlesztett minION készülékek, a másik pedig a a PacBio RSII és Sequel

készülékei. A két technológia különböző szekvenálási elveken alapszik. A PacBio szekvenálások során a megszekvenálandó molekula egy hajtű alakú adapterhez van ligálva, melyet egy SMRTcellre feltöltve, egy felülethez rögzített polimeráz fog lemásolni fluoreszcensen jelölt nukleotidokkal, melynek a fluoreszcenciáját egy optikai szenzor fogja detektálni¹⁴. Ezzel szemben az ONT-minION-nál a szekvenálás egy nanopóruson fog megtörténi, melyet lipid kettősrétegbe ágyazott alfa-hemolizinek fognak alkotni^{15,16}. A megszekvenálandó molekulához ezen könyvtár készítés esetében egy motorfehérjét ligálnak, amely az RNS molekula bejutását segíti a pórusba^{7,17}. A pórusba bejutott RNS molekula egy feszültségeltolódást fog létrehozni, amely egy adott szekvenciára jellemző. Ezenkívül az ONT-minION használatával lehetőség nyílik direkt RNS szekvenálások kivitelezésére is¹⁸. Nagy előnye, hogy a könyvtár készítés során nincs szükség sem reverz transzkripcióra, sem PCR amplifikációra, ezáltal kizárólag csak a natív RNS molekula kerül megszekvenálásra. Alkalmazásával hatékonyan kiküszöbölhetőek olyan hibák a transzkriptom analízisben mint a templát szál

váltás jelensége és a false priming, melyek hamis intron és transzkripció végpont helyek detekcióját eredményezhetik¹⁹. Ezen két platform megjelenésével a transzkriptomról alkotott képünk sokkal komplexebbé vált az elmúlt évek során. Egyik érdekes alkalmazási területük közé tartozik a virális transzkriptomok analízise, használatukkal az újonnan felfedezett transzkript izoformák, nem kódoló RNS-ek és intron variánsok száma megsokszorozódott és a virális transzkriptom szabályozásainak új aspektusait világította meg²⁰.

Az **African Swine Fever (ASFV)** egy nukleocitoplazmatikus (NCLDV) vírus, amely az Asfarviridae családba tartozik²¹. Sertéseket és vaddisznókat fertőz, akut, halálos vérzéses betegséget okozva. Replikációja a gazdaszervezet monocita/makrofág sejtjeiben fog megtörténni²². Ikozaéderes morfológiájú, átmérője ~200 nm^{23,24}. Genomja lineáris kettősszálú DNS, melynek mérete 170-190 kb. Az osztódáshoz és mRNS szintézishez szükséges minden fehérjét tartalmazni fogja a virion, ezért replikációja a citoplazmában fog történni^{25,26}.

Genomjának szerkezeti felépítésére jellemző, hogy közepén tartalmaz egy konzervált és egy variábilis régiót és a két szélén egy inverted terminális repeatet, ahol a vírus replikációja fog megtörténni a Poxvírusokhoz hasonlóan^{25,27}. A vírus mRNS-ei 5' cap és 3' poly(A) farokkal rendelkeznek^{28,29}. RNS-einek szintézisét a Poxvírusokra is jellemző temporális génexpresszió jellemzi^{30,31}. RNS-einek szintézise négy kinetikai osztályba sorolható: azonnali-korai, korai, köztes és késői^{32,33}. Az azonnali korai és korai gének a replikációban szerepet játszó fehérjék kódolásáért lesznek felelősek, míg a köztes és késői gének a vírus szerkezeti fehérjeit fogják kódolni. A vírus mRNS-eit korábban Cackett és munkatársai CAGE és poly(A) rövid read szekvenálásokkal már vizsgálta³⁴.

A Pseudorabies virus (PRV) egy sertészherpesz vírus, amely az alfaherpeszek családjába tartozik³⁵. Sertésekben idegrendszeri megbetegedéseket okoz^{36,37}. Genomja kettősszálú lineáris DNS, mely ~143 kb hosszú³⁸⁻⁴⁰. Szerkezetileg a vírus DNS-e két részből épül fel az UL (long unique) és US (short unique) régióból⁴¹.

Az US régiót egy inverted és egy terminális repeat határolja, amelyek remkombinációja a genom kétféle izomerét eredményezheti. Ezenkívül az IE180 és US1 gének kettő kópiában találhatóak ebben a régióban⁴². Az UL régió egy replikációs origót az OriL-t, míg az US régió kettő replikációs origót az OriS-et fogja tartalmazni^{43,44}. A PRV gégeinek többsége policisztronos formában fog kifejeződni⁴⁵⁻⁴⁷. Transzkripciójához a gazdaszervezet transzkripció apparátusát fogja használni, de néhány transzkripció faktort, mint például az IE180, US1 és EP0-át a vírus fogja kifejezni⁴². Gégei három kinetikai osztályba sorolhatóak: azonnali-korai, korai, és késői⁴⁸. A fertőzés során legelőször az immediate-early gégei fognak expresszálni, amely transzkripció aktivátorokat, úgy mint IE180-at fog kifejezni⁴². Ezután az early gégek fejeződnék ki, amelyek a replikációban fognak szerepet játszani. Replikációja gördülő gyűrű mechanizmussal történik⁴⁹. Az early gégek kifejeződését a late gégek fogják követni, amelyek strukturális proteinek kódolnak^{39,42}. A vírus a fertőzés után a gazdaszervezet trigeminális ganglionjában fog jelen maradni^{42,50}. Ebben a látenciában a LAT (Látencia asszociált transzkript) gén az

egyetlen, mely transzkripciósan aktív. Ha az IE180 fehérje hozzákötődik, akkor expressziója gátolt lesz. Ennek érdekében, hogy a vírus jelen maradhasson a gazdasejtben, a LAT expressziója segíti a neuronok túlélését és anti-apoptótikus hatásokat.

Célkítűzések

Tézisemben az African Swine Fever és Pseudorabies vírusok transzkriptom analízisét mutatom be. Ezen vizsgálatokhoz hosszú leolvasásokat adó szekvenáló platformokat, úgy mint PacBIO és ONT-minION készülékeket alkalmaztunk.

A következő célokat tűztük ki az analízis során:

1. Mindkét vírus esetében a virális mRNS-ek transzkripcióos kezdőhelyeinek és transzkripcióos végpontjainak a meghatározása bázispár pontossággal;
2. Ezen annotált transzkripcióos kezdőhelyek és transzkripcióos végpontok összekötése teljes hosszúságú transzkriptekké;

3. Az annotált transzkriptek strukturális kategorizálása, úgy mint az egyes izoformák, nem kódoló RNS-ek, antiszensz, vagy hipotetikus transzkriptjeinek a meghatározása;
4. Promóter elemek és poliadenilációs szignálok azonosítása az RNS-ek 5' és 3' végeinek esetében;

Anyagok és módszerek

Az African Swine Fever vírust primer alveoláris makrofág (PAMs) sejteken, míg a Pseudorabies vírust porcine kidney-15 (PK-15) sejteken tenyésztettük. A fertőzéseket különböző időpontokban leállítottuk, majd RNS-t izoláltunk belőlük. Mindkét vírus totál RNS-én poly(A) szelekciót végeztünk, a PRV esetében a minták egy részén történt ribodepléció és cap szelekció is. A poly(A) szelektált mintákból PCR amplifikált cDNS és nem amplifikált dcDNS, valamint nativ dRNS szekvenáló ONT-minION könyvtárak készültek. A PRV esetében PacBio RSII és Sequel szekvenáló könyvtárakat is készítettünk. Ezt követően ezen LRS platformokon megszekvenáltuk a virális transzkriptomot. A szekvenátorok által létrehozott adatokat basecolloltuk, a

PacBio esetében SMRTlink programcsomaggal, míg az ONT esetében a Guppy szofverrel. A readeket a referencia genomhoz térképeztük a Minimap2 program segítségével. Ezt követően a keletkezett fájlokban lévő transzkriptek kvantitálására és jellemzésére egy python alapú transzkript annotáló LoRTIA programcsomagot használtunk. A leolvasások és az annotált transzkriptek ábrázolásához IGV és Genious szoftvereket alkalmaztunk.

Eredmények

A PRV esetében 465 transzkripció kezdőhelyet és 57 transzkripció végpontot azonosítottunk, melyekből összesen 619 transzkriptet annotáltunk. Ezekből 166 hosszú és 24 rövid 5' UTR izoformát, valamint 22 3' UTR variánst határoztunk meg. Emellett 209 5' truncated transzkriptet is azonosítottunk. Ezek olyan RNS-ek, amelyek génen belüli in-frame ORF-et tartalmaznak. Az LRS szekvenálások egyik nagy felfedezése a replikáció asszociált transzkriptek voltak. Ezekből az OriL esetében a CTO transzkriptek új hossz és alternatív terminációs izoformáit detektáltuk, valamint az US régióban lévő OriS PTO-US1 új splicolt izoformáját is azonosítottuk.

Ezenkívül azonosítottunk hosszú nem kódoló transzkript (>200bp) izoformákat, melyek az US régióban lokalizálódnak. Ilyen volt például az AST és LLT transzkriptekkel koterminálsan végződő NOIR transzkript új hossz és splice izoformái, valamint az US3 génnel antiszensz állású Azure transzkriptek. Az előzőleg már annotált US4 antiszensz transzkript TSS-ét is sikerült pontosítanunk. Emellett az UL15 intron régiójából induló in-frame ATG-t nem tartalmazó fORF15 transzkript új hossz variánást is azonosítottuk. A hosszú leolvasások eredményeként több gént átíró úgynevezett multicisztronos transzkripteket is azonosítottunk. Ezekből 87 policisztronos volt, míg 24 darab komplex transzkript.

Az ASFV vírusban 202 TSS-t és 220 TES-t azonosítottunk, melyekből összesen 311 transzkriptet annotáltunk. Ezekből 14 hosszú és 2 rövid 5' UTR izoformát, valamint 57 3' UTR különítettünk el. 5' truncated transzkriptekből 16 db-ot azonosítottunk. Ezek mellett találtunk 7 hipotetikus gént is, melyek kis ORF-eket tartalmaznak a vírus intergenikus régiójában. Több nem kódoló RNS-t is azonosítottunk, úgy mint 3' csonkolt

RNS-t, antiszensz és replikáció asszociált RNS-eket. A 3' csonkolt RNS-ekből összesen 22 db-ot azonosítottunk, ezen RNS-ek tartalmazzák a kanonikus ORF-ek ATG-it, de nincs bennük STOP kodon, vagyis nincs funkcióképes ORF-ük. Antiszensz RNS-ekből összesen 7 darabot azonosítottunk. Ezenkívül a genom végeken található inverted repeatekben, ahonnan a replikáció elindul, találtunk 7 darab alacsony abundanciájú replikáció asszociált RNS-t. A közelmúltig, az volt a feltételezés, hogy az ASFV vírus csak monocisztronos RNS-eket fejez ki^{33,51}. Az LRS által fény derült arra, hogy multicisztronos RNS-eket is kifejez. Policisztronos RNS-ekből 51 darabot, míg komplex RNS-ekből 22 darabot annotáltunk.

Diszkusszió

Tézisemben az ASFV és PRV transzkriptomának vizsgálatát tárgyalom hosszú leolvasásokat adó szekvenáló platformok alkalmazásával. Ezen megközelítéseket alkalmazva számos új transzkript kategóriát azonosítottunk, amelyek eddig ismeretlenek voltak.

Vizsgálataink eredményeként számos új transzkripciós kezdőhelyet és transzkripciós végpontot azonosítottunk. Ezenkívül új 5' és 3' UTR izoformákat fedeztünk fel. Az 5' UTR-ek upstream ORF-eket tartalmazhatnak, amelyek a transláció hatékonyságát befolyásolhatják⁵². A 3' UTR-ek olyan szekvenciákat tartalmazhatnak, amelyek mikroRNS-eket köthetnek, befolyásolva ezáltal a fehérjék szintézisét⁵³.

Az LRS szekvenálások az 5' csonkolt hipotetikus RNS-ek széles spektrumát tárták fel. Ezek olyan RNS-ek, amelyek génen belüli ORF-ekről íródnak le, vagyis a kanonikus génnel közös STOP kodont tartalmaznak, de az ATG-jük a kanonikus ATG-től downstream található meg. Ezenkívül nemcsak génen belüli 5' csonkolt RNS-eket azonosítottunk, hanem az ASFV esetében az intergenikus régiókban kis ORF-et tartalmazó RNS-eket is találtunk. A nem kódoló RNS-ek között azonosítottunk még antiszensz transzkripteket, melyek átíródhatnak önálló promóterek által (LLT, AST, AZURE), vagy létrejöhetnek konvergens génpárok közötti transzkripciós átírások vagy divergens génpárok 5' UTR-jeinek átfedése által⁵⁴. Számos ilyen

antiszensz transzkriptet azonosítottunk PRV-ben és ASFV-ben. Erre példa PRV esetében az US régióban található Azure, amely az US3-al antiszensz, valamint az US4-AS, amely az US4 génnel áll ellentétes orientációban. ASFV esetében 7 darab ilyen antiszensz állású molekulát detektáltunk.

Ezen platformok egy másik jelentős eredménye volt, hogy, az RNS-ek egy új osztályát sikerült velük azonosítani, melyek a replikációs origót átíró és a replikációs origók közelében lévő transzkripteket foglalják magukba⁵⁵. Őket replikáció asszociált RNS-eknek (raRNA) nevezzük. A PRV esetében, mindhárom replikációs origó körül annotáltunk ilyen molekulákat. Analízisünk feltárta, hogy a legabundánsabban expresszálandó RNS a PRV UL régiójában található OriL melletti CTO-S molekula. A CTO-S nek azonosítottunk Ori-t átfedő mRNS izoformáit, valamint alternatív terminációs izoformáit is, mely a szomszédos UL22 génnel fed át. Ezen RNS-ek biológia szerepe tisztázatlan, valószínűleg a replikációs beindításában és a replikációs villa irányának a meghatározásában lehet szerepük,

ugyanakkor poly(A) szekvenciákkal is rendelkeznek, ami arra utal, hogy mRNS molekulaként is működhetnek a vírus élelciklusában⁵⁵. Ezenkívül más funkciójuk is lehet, mint például RNS-DNS hibrideket képeznek⁵⁶. A vírus US régiójában található OriS körül is azonosítottunk ilyen molekulákat. Ezek közül a PTO-US1 új splice variánsát, amely átéri a replikációs origót és az US1-el végződik közös koterminálisban, valamint azonosítottuk a NOIR-1 transzkriptek splice és hossz izoformáit, melyek az OriS közelében lokalizálódnak és az LLT/AST génekkel van közös 3' koterminálisuk. Ezen átfedések lehetséges szerepe az lehet, hogy a NOIR-1 valamilyen módon gátolja ezen látencia gének kifejeződését. Azonban, ezen hipotézis bizonyítása további vizsgálatokat igényel. ASFV esetében is azonosítottunk 6 ilyen raRNAs molekulát, melyek a genom végeken található inverted repeatekben íródnak le, ahonnan a replikáció elindul. Közülük azonosítottunk nem kódoló transzkripteket, valamint a fehérjét kódoló génből induló mRNS izoformát, amelyek a terminális repeatekben lévő nem kódoló transzkriptekkel végződnek koterminálisan.

A multicisztronos RNS-ek több gént átfedő molekulák, melyek átírhatnak azonos polaritású géneket (policisztronos transzkript) és ellentétes polaritású géneket (komplex transzkript). Analízisünkben, mind a PRV és mind az ASFV vírusokban számos ilyen hosszú RNS molekulát detektáltunk.

Az eredményeink alapján azonosított 5' és 3' UTR átfedések, valamint a multicisztronos transzkriptek, a transzkripciós átfedések egy széles hálózatát hozzá létre a virális genomon. Az átfedések orientáció alapján lehetnek divergenssek, konvergensek és paralelelek. Feltételezzük, hogy lehetséges szerepük a transzkripciós interferenciában lehet, amely a szomszédos gének kifejeződésére gyakorol szabályozó hatást a virális genomon létrejövő transzkripciós átfedések által⁵⁷.

Hivatkozásjegyzék

1. K-H Liang. Bioinformatics for Biomedical Science and Clinical Applications - 1st Edition.

<https://shop.elsevier.com/books/bioinformatics-for->

- biomedical-science-and-clinical-applications/liang/978-1-907568-44-2 (2023).
2. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
 3. Heather D. VanGuilder. Twenty-five years of quantitative PCR for gene expression analysis | BioTechniques. <https://www.future-science.com/doi/10.2144/000112776> (2023).
 4. Southern, E. The early days of blotting. *Methods Mol Biol* **1312**, 1–3 (2015).
 5. Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016).

6. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: Lessons from Large-Scale Biology. *Science* **300**, 286–290 (2003).
7. Dijk, E. L. van, Jaszczyszyn, Y., Naquin, D. & Thermes, C. The Third Revolution in Sequencing Technology. *Trends in Genetics* **34**, 666–681 (2018).
8. Metzker, M. L. Sequencing technologies — the next generation. *Nat Rev Genet* **11**, 31–46 (2010).
9. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364 (2012).
10. Weimer, B. C. 100K Pathogen Genome Project. *Genome Announc* **5**, e00594-17 (2017).
11. Turnbull, C. *et al.* The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ* **361**, k1687 (2018).

12. van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. Ten years of next-generation sequencing technology. *Trends Genet* **30**, 418–426 (2014).
13. Byrne, A., Cole, C., Volden, R. & Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20190097 (2019).
14. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
15. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* **14**, 265–279 (2016).
16. Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nat Biotechnol* **30**, 326–328 (2012).

17. Satam, H. *et al.* Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **12**, 997 (2023).
18. Garalde, D. R. *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* **15**, 201–206 (2018).
19. Cocquet, J., Chong, A., Zhang, G. & Veitia, R. A. Reverse transcriptase template switching and false alternative transcripts. *Genomics* **88**, 127–131 (2006).
20. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research. *Trends in Microbiology* **27**, 578–592 (2019).
21. Reis, A. L., Netherton, C. & Dixon, L. K. Unraveling the Armor of a Killer: Evasion of Host Defenses by

- African Swine Fever Virus. *J Virol* **91**, e02338-16 (2017).
22. Zhu, J. J. *et al.* Mechanisms of African swine fever virus pathogenesis and immune evasion inferred from gene expression changes in infected swine macrophages. *PLOS ONE* **14**, e0223955 (2019).
23. Andrés, G., Simón-Mateo, C. & Viñuela, E. Assembly of African swine fever virus: role of polyprotein pp220. *J Virol* **71**, 2331–2341 (1997).
24. Salas, M. L. AFRICAN SWINE FEVER VIRUS (ASFARVIRIDAE). in *Encyclopedia of Virology (Second Edition)* (eds. Granoff, A. & Webster, R. G.) 30–38 (Elsevier, 1999). doi:10.1006/rwvi.1999.0008.
25. Dixon, L. K., Chapman, D. A. G., Netherton, C. L. & Upton, C. African swine fever virus replication and genomics. *Virus Research* **173**, 3–14 (2013).

26. Alejo, A., Matamoros, T., Guerra, M. & Andrés, G.
A Proteomic Atlas of the African Swine Fever Virus Particle. *J Virol* **92**, e01293-18 (2018).
27. Yoo, D., Kim, H., Lee, J. Y. & Yoo, H. S. African swine fever: Etiology, epidemiological status in Korea, and perspective on control. *Journal of Veterinary Science* **21**, (2020).
28. Du, X., Gao, Z.-Q., Geng, Z., Dong, Y.-H. & Zhang, H. Structure and Biochemical Characteristics of the Methyltransferase Domain of RNA Capping Enzyme from African Swine Fever Virus. *J Virol* **95**, e02029-20 (2021).
29. Salas, M. L., Kuznar, J. & Viñuela, E.
Polyadenylation, methylation, and capping of the RNA synthesized in vitro by African swine fever virus. *Virology* **113**, 484–491 (1981).

30. Gaudreault, N. N., Madden, D. W., Wilson, W. C., Trujillo, J. D. & Richt, J. A. African Swine Fever Virus: An Emerging DNA Arbovirus. *Front. Vet. Sci.* **7**, 215 (2020).
31. Wang, Y. *et al.* Structure of African Swine Fever Virus and Associated Molecular Mechanisms Underlying Infection and Immunosuppression: A Review. *Front. Immunol.* **12**, 715582 (2021).
32. Cackett, G., Portugal, R., Matelska, D., Dixon, L. & Werner, F. African Swine Fever Virus and Host Response: Transcriptome Profiling of the Georgia 2007/1 Strain and Porcine Macrophages. *Journal of Virology* **96**, (2022).
33. Rodríguez, J. M. & Salas, M. L. African swine fever virus transcription. *Virus Research* **173**, 15–28 (2013).

34. Cackett, G. *et al.* The African Swine Fever Virus Transcriptome. *J Virol* **94**, e00119-20 (2020).
35. Pellett, P. E. *et al.* Order - Herpesvirales. in 99 (Elsevier, 2012). doi:10.1016/B978-0-12-384684-6.00005-7.
36. Xiang, S. *et al.* Complete Genome Sequence of a Variant Pseudorabies Virus Strain Isolated in Central China. *Genome Announcements* **4**, 10.1128/genomea.00149-16 (2016).
37. Zheng, H.-H. *et al.* Seroprevalence investigation and genetic analysis of pseudorabies virus within pig populations in Henan province of China during 2018–2019. *Infection, Genetics and Evolution* **92**, 104835 (2021).

38. Zheng, H.-H., Fu, P.-F., Chen, H.-Y. & Wang, Z.-Y. Pseudorabies Virus: From Pathogenesis to Prevention Strategies. *Viruses* **14**, 1638 (2022).
39. Nauwynck, H., Glorieux, S., Favoreel, H. & Pensaert, M. Cell biological and molecular characteristics of pseudorabies virus infections in cell cultures and in pigs with emphasis on the respiratory tract. *Vet. Res.* **38**, 229–241 (2007).
40. Zhou, M. *et al.* Characterization of a moderately pathogenic pseudorabies virus variant isolated in China, 2014. *Infection, Genetics and Evolution* **68**, 161–171 (2019).
41. Klupp, B. G., Hengartner, C. J., Mettenleiter, T. C. & Enquist, L. W. Complete, Annotated Sequence of the Pseudorabies Virus Genome. *Journal of Virology* **78**, 424–440 (2004).

42. Pomeranz, L. E., Reynolds, A. E. & Hengartner, C. J. Molecular Biology of Pseudorabies Virus: Impact on Neurovirology and Veterinary Medicine. *Microbiol Mol Biol Rev* **69**, 462–500 (2005).
43. Fuchs, W., Ehrlich, C., Klupp, B. G. & Mettenleiter, T. C. Characterization of the replication origin (OriS) and adjoining parts of the inverted repeat sequences of the pseudorabies virus genome. *Journal of General Virology* **81**, 1539–1543 (2000).
44. Klupp, B. G., Kern, H. & Mettenleiter, T. C. The virulence-determining genomic BamHI fragment 4 of pseudorabies virus contains genes corresponding to the UL15 (partial), UL18, UL19, UL20, and UL21 genes of herpes simplex virus and a putative origin of replication. *Virology* **191**, 900–908 (1992).

45. Mainguy, G., Koster, J., Woltering, J., Jansen, H. & Durston, A. Extensive Polycistronism and Antisense Transcription in the Mammalian Hox Clusters. *PLOS ONE* **2**, e356 (2007).
46. Tombácz, D. *et al.* Transcriptome-wide survey of pseudorabies virus using next- and third-generation sequencing platforms. *Sci Data* **5**, 180119 (2018).
47. Moldován, N. *et al.* Multi-Platform Sequencing Approach Reveals a Novel Transcriptome Profile in Pseudorabies Virus. *Frontiers in Microbiology* **8**, (2018).
48. Honess, R. W. & Roizman, B. Regulation of Herpesvirus Macromolecular Synthesis I. Cascade Regulation of the Synthesis of Three Groups of Viral Proteins. *Journal of Virology* **14**, 8–19 (1974).

49. Berthomme, H., Monahan, S. J., Parris, D. S., Jacquemont, B. & Epstein, A. L. Cloning, sequencing, and functional characterization of the two subunits of the pseudorabies virus DNA polymerase holoenzyme: evidence for specificity of interaction. *J Virol* **69**, 2811–2818 (1995).
50. Deng, J., Wu, Z., Liu, J., Ji, Q. & Ju, C. The Role of Latency-Associated Transcripts in the Latent Infection of Pseudorabies Virus. *Viruses* **14**, 1379 (2022).
51. Almazán, F., Rodríguez, J. M., Angulo, A., Viñuela, E. & Rodriguez, J. F. Transcriptional mapping of a late gene coding for the p12 attachment protein of African swine fever virus. *Journal of Virology* **67**, 553–556 (1993).

52. Kronstad, L. M., Brulois, K. F., Jung, J. U. & Glaunsinger, B. A. Reinitiation after translation of two upstream open reading frames (ORF) governs expression of the ORF35-37 Kaposi's sarcoma-associated herpesvirus polycistronic mRNA. *J Virol* **88**, 6512–6518 (2014).
53. Cackett, G., Sýkora, M. & Werner, F. Transcriptome view of a killer: African swine fever virus. *Biochemical Society Transactions* **48**, 1569–1581 (2020).
54. Torma, G. *et al.* An Integrated Sequencing Approach for Updating the Pseudorabies Virus Transcriptome. *Pathogens* **10**, 242 (2021).
55. Boldogkői, Z., Balázs, Z., Moldován, N., Prazsák, I. & Tombácz, D. Novel classes of replication-

- associated transcripts discovered in viruses. *RNA Biol* **16**, 166–175 (2019).
56. 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).
57. Boldogkoi, Z. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Frontiers in Genetics* **3**, (2012).