

VÍRUSSTRANSZKRIPTOMOK ÁTFOGÓ KARAKTERIZÁLÁSA HOSSZÚ LEOLVASÁSOKAT ADÓ SEKVENÁLÁS SEGÍTSÉGÉVEL

PhD tézisfüzet

Moldován Norbert



Orvosi Biológiai Intézet
Interdiszciplináris Orvostudományok
Doktori Iskola
Általános Orvostudományi Kar
Szegedi Tudományegyetem
Témavezető: Prof. Boldogkői Zsolt, PhD, DSc

Szeged
- 2020 -

A tézishez közvetlenül kapcsolódó publikációk listája

- I. Moldován, N. et al. (2018) ‘Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus’, Scientific Reports. doi: 10.1038/s41598-018-26955-8.
- II. Boldogkői, Z. et al. (2019) ‘Long-Read Sequencing – A Powerful Tool in Viral Transcriptome Research’, Trends in Microbiology, pp. 578–592. doi: 10.1016/j.tim.2019.01.010.
- III. Tombácz, D. et al. (2019) ‘Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome’, Frontiers in genetics, 10, p. 834.

Bevezető

Egy élőlény fenotípusát génjeinek expressziója határozza meg, melyet a környezet befolyásol. Crick a biológia centrális dogmájában mutatta be először az információ áramlását a géntől a fehérjéig, melyben a hírvívő az RNS (F. H. Crick 1958; F. Crick 1970). Egy adott időpillanatban a transzkriptom, azaz az RNS molekulák összessége, jellemzi a sejtet, melyben ezek kifejeződtek. Az RNS-seq egy elterjedt eszközzé vált a génexpresszió kísérleti csoportok közötti kvantitatív elemzésére (ez az ún. differenciál-génexpresszió elemzés, vagy DGE) (Young et al. 2012), ugyanakkor lehetővé teszi mikroorganizmus és szövet transzkriptomok longitudinális vizsgálatát is (Hubbard et al. 2013). Mindemellett alkalmas a transzkriptek strukturális jellemzésére, így vizsgálhatóvá válnak az alternatív splicing események (E. T. Wang et al. 2008), és hossz izoformák (Depledge et al. 2019). Az RNS-seq által vált világossá a nem kódoló RNS-ek génexpressziót szabályozó szerepe (Djebali et al. 2012; Morris and Mattick 2014), illetve az RNS nukleotid módosulások RNS érésben és expresszió szabályozásban betöltött funkciója (Schaefer, Kapoor, and Jantsch 2017; Liu et al. 2019).

Az első RNS szekvenálásokat a Sanger-féle szekvenálás hajnalán hajtották végre (M. Adams et al. 1991; M. D. Adams et al. 1995). Ez az ún. első generációs szekvenálás olyan technikai nehézségektől szenvedett, mint a túl rövid leolvasások, vagy az alacsony áteresztőképesség. A nanotechnológia és a mikrofluidika területén történő áttörések azonban egy új

szekvenáló platform generációjának lehetőségét vetítették előre. A szekvenátorok következő generációja milliányi cDNS molekula szekvenálására képesek párhuzamosan, így csökkentve a szekvenálás költségeit és időtartamát, és ugrásszerűen növelve az eredményül kapott információ mennyiségét. Második generációs technológiákkal már vizsgálhatóvá váltak a splice-variánsok kifejeződési szintjei, és már alkalmasak új gének és nem kódoló RNS-ek detektálására is (Z. Wang, Gerstein, and Snyder 2009). Jelenleg azonban épp a szekvenálás harmadik forradalmának közepén vagyunk. A harmadik generációs szekvenátorok képesek akár nagyon hosszú leolvasásokat produkálni, miközben megőrzik nagy áteresztőképességüket. A Pacific Biosciences egy már korábban alkalmazott fényfelvillanásos módszerre alapozta harmadik generációs technológiáját, mely alkalmas egyetlen molekula megszekvenálására is. A piac új versenyzője az Oxford Nanopore Technologies cég, mely a szekvenálást egészen új megközelítést valósította meg. Szekvenátoruk szintetikus membránba ágyazott protein pórusokon és az áramerősség mérésén alapul, melyet a pórusokon áthaladó nukleotidok befolyásolnak.

A szekvenálást az RNS kivonása, tisztítása és a könyvtárkészítés előzi meg. Minden lépés alkalmas műtermékek létrehozására, melyeket az adatok előfeldolgozásának kell szűrnie. A végső analízisnek pedig óriási adathalmazból kell értelmezhető eredményt produkálnia.

A HSV-1: Az 1-es típusú herpes simplex vírus az *Alphaherpesvirinae* alcsaládba tartozó herpeszvírus. Az emberi populáció egyik legelterjedtebb vírusa, mely az ajkak körüli fájdalmas hójagokat okozza, súlyos esetekben pedig agyveőgyulladásához vezethet. Gazdaszervezetében élethosszig tartó látenciát alakíthat ki. DNS genomja 152 kbp hosszú, lineáris, kétszálú, és egy ikozahedrális kapszidba csomagolt, melyet lipidburok borít. A genomot egy hosszú (UL) és egy rövid (US) alegység építi fel, mindkettőt inverz ismétlődések határolják (IRL és IRS) (Macdonald et al. 2012). 72 génje által kifejezett transzkriptjeit rövid leolvasásokat adó szekvenálással vizsgálták korábban (Rajcáni, Andrea, and Ingeborg 2004).

A BoHV-1: A HSV-1 nem humán patogén rokona a szarvasmarhák 1-es típusú alfaherpeszvírusa (BoHV-1), mely világszerte fertőz. Légúti és termékenységgel kapcsolt tüneteket okoz, és a szarvasmarhák fertőző rhinotracheitis-ének (IBR) legfőbb forrása (Muyllkens et al. 2007). Kétszálú DNS genomja 135 kbp hosszú, és magas, 72%-os GC tartalom jellemzi. Felépítésében két egyedi szekvencia, az UL és az US vesz részt, utóbbi inverz ismétlődő szekvenciák keretezik (d'Offay, Fulton, and Eberle 2013). A vírus 73 génje más alfaherpeszvírusokéval nagyjából homológ, a gének nomenklatúrája a HSV-1-ét követi. Ez alól két kivétel a circ gén, mely a *Varicellovirus* génuszra jellemző, és az ul0.5 mely a BoHV-1 sajátja (Delhon et al. 2003). A génexpresszió kinetikája az alfaherpeszvírusokéra jellemző, három kategóriába sorolható: azonnali-korai,

korai és kései génekre. Az azonnali-korai gének a transzkripcióban, a korai gének a replikációban játszanak szerepet, míg a késeiek a vírus struktúrális fehérjéit kódolják.

Az AcMNPV: Az *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) a *Baculoviridae* családból származó rovarvírus (Blissard and Rohrmann 1990). A biotechnológiában fehérjeexpressziós rendszerek recombináns vektoraként, a mezőgazdaságban pedig biológiai rovarölő ágensként használatos (Hu 2005). Kétszálú DNS genomja körkörös, 133 kbp hosszú, és 156 szorosán elhelyezett gént tartalmaz. A gén expressziója három fázisba sorolható: korai, kései és nagyon kései. A korai gének promótere TATA box, és a gazda transzkripciós apparátusának szolgál szubsztrátul. A korai gének transzkripciója a rovarok között elterjedt CAGT iniciátorról indul (a +1 nt aláhúzva) (Kogan et al. 1995). A kései és nagyon kései génjeit a vírus saját RNS polimeráza írja át, egy konzervált TAAG szekvencia motívumtól kezdve, melyet kései iniciátor szekvenciának neveznek (Garrity, Chang, and Blissard 1997).

Az epitarnszkriptomika egy dinamikusan fejlődő tudományterület. A cap struktúra 7-metilguaninján (m^7G) kívül az RNS több mint 60 egyéb módosulást tartalmaz (Zhao, Roundtree, and He 2018), többek között más metilált bázisokat, izomerizált pseudouridint, vagy metilált bázisok oxidált formáit (Roundtree et al. 2017). Korábbi metilációs vizsgálatok a simian vírus 40-es típusában (Lavi and Shatkin 1975), adenovírusokba

(Sommer et al. 1976) és a HSV-1-ben (Moss et al. 1977) is találtak RNS metilációt. Az ismert nukleotid módosulások számának ellenére funkciójuk nagyrészt ismeretlen. A 6-metiladenozin replikációban és a HIV-1 vírus-gazda interakciójában betöltött szerepéről több korábbi kutatás beszámolt (Lichinchi et al. 2016; Tirumuru et al. 2016). A vírus RNS-ek metilációját a sejt (Gokhale et al. 2016) és a vírus saját metiltranszferázai végzik (Tao et al. 2013; Ho, Gong, and Shuman 2001). Az AcMNPV epitranszkriptomával még nem foglalkozott korábbi tanulmány.

Célkitűzések

Tézisemben bemutatom a hosszú leolvasásokat adó szekvenálási technikák alkalmazási lehetőségeit a vírus transzkriptom vizsgálatában. Ehhez csoportunk által tanulmányozott négy vírust használok. A HSV-1 segítségével a harmadik generációs technológiák izoforma annotációban betöltött szerepét mutatom be. A BoHV-1 segítségével a virális génexpressió feltárásának lehetőségeit prezentálom, míg az AcMNPV transzkriptjei esetén az RNS módosulások detektálásának lehetőségeit tárgyalom. Ezen fókuszpontokon kívül az egyes vírusok analízise során kapott eredményeket más pontokba is bemelelem, ahol szükséges.

Anyagok és módszerek

A vírusok szaporítása a megfelelő sejtvonalakban történt, a sejtvonalak gyártói által ajánlott körülmények között. A fertőzések több egymást követő időpontban lettek állítva, majd megtörtént a totál RNS izolálása. A HSV-1 és BoHV-1 minták egy részén rRNS depléción történt, a HSV-1 és AcMNPV minták egy részén pedig cap-szelekció. Mindhárom vírus mintáinak egy részén poly(A)-szelekció is történt. A HSV-1 és az AcMNPV mintákból PacBio IsoSeq könyvtárak, míg mindhárom vírus mintáiból amplifikált teljes hosszúságú cDNS könyvtárak készültek. A BoHV-1 poly(A) szelektált mintából nem amplifikált cDNS könyvtárakat is létrehoztunk a longitudinális expresszió vizsgálatához. A könyvtárakat PacBio és MinION platformokon szekvenáltuk. A keletkező adatokat bázishívtuk, előfeldolgoztuk, térképeztük, majd a LoRTIA szoftvercsomaggal megtörtént az izoformák meghatározása. A BoHV-1 génexpresszió analíziséhez a leolvasásokat a DESeq2 szoftvercsomag Medián Adag Normalizálásával normalizáltuk (Wu et al. 2019). A nukleotid módosulásokat a Tombo szoftvercsomaggal detektáltuk (Stoiber et al. 2016).

Eredmények

A vírus transzkriptom annotálása: 182 új feltételezeten protein kódoló mRNS-t, 8 ncRNS-t, 53 új transzkripciósi starthely (TSH) és alternatív terminációjú RNS-t detektáltunk a HSV-1 transzkriptomában. A harmadik generációs technológia képes a splice izoformák pontos detektálására is. Többek közt az UL34-35 és az RL1-RL2 splice izoformáit is megtaláltuk. 201 multigénes transzkriptet azonosítottunk, melyek közül sok policisztronikus, míg mások úgynevezett komplex transzkriptek, melyek a transzkript irányával ellentétes nyitott leolvasási kereteket is tartalmaznak.

A sok multigénes illetve hosszú TSH és alternatívan végződő transzkript átfedéseket eredményez a HSV-1 transzkriptomában. Gyakorlatilag az összes konvergensen végződő gén között találtunk átfedést, melyet az RNAPII továbbhaladása eredményez az mRNS 3' végi hasítását követően (Proudfoot, 2016). Ezek a molekulák azonban rövid életűek, mivel az exonukleázok megemésztik őket.

A virális génexpresszió analízise: A génexpresszió elemzését a BoHV-1 nem amplifikált cDNS könyvtárán végeztük. A TATA box-szal rendelkező TSH-k abundanciája az időben gyorsabban növekszik, mint azoké, amelyek nem rendelkeznek TATA box-szal, utóbbiak a fertőzést követő 4. órában érik el csúcspontjukat. Ugyanez igaz a poly(A) szignállal rendelkező terminációk abundanciája is, mely gyorsabban növekszik a fertőzés ideje alatt, mint azoké,

amelyek esetén hiányzik a poly(A) szignál. Előbbiek abundanciája a fertőzést követő 12 órában háromszorosan meghaladja utóbiakét. A BoHV-1 génexpressziós mintázata az alfaherpeszvírusokét követi. Génjeit legalább három csoportba tudtuk osztani: azonnali-korai, korai és kései gének. Az első órában a vírus transzkripciójáért felelős gének már kifejeződnek (bicp0, bicp22). Azt találtuk, hogy a CIRC, mely egy ismeretlen funkciójú mirisztilált tegument fehérje mRNS-e (Fraefel, Ackermann and Schwyzer, 1994) az első órában szintén megjelenik, mely a fehérje transzkripcióban, vagy replikációban játszott szerepére utalhat. Pokhriyal és mtsai. megfigyelték, hogy három korábban késeinek hitt gén (ul21, ul33 és ul34) a vírusfertőzést követően közvetlenül kifejeződnek (Pokhriyal *et al.*, 2018). Eredményeink azt mutatják, hogy az ul33 és az ul34 korai expressziós mintázatot mutatnak. Negyvenegy másik gén transzkripció aktivitása is detektálható a fertőzés második órájában, de 69 génből 66 a maximum abundanciáját csak a replikáció beálltát követően éri el. Huszonhárom gén, többek között az ul12, mely az újonnan szintetizált DNS további feldolgozásáért felelős, illetve több tegumentum és kapszid fehérjét kódoló gén a fertőzést követő negyedik órában mérhető expressziót mutat, expresszója pedig a vírusfertőzés ideje alatt emelkedik. Ezek a kései gének első hullámát képviselik, és köztük találhatóak a legabundánsabb transzkriptek. A DNS replikációját követő időszakban mindössze négy gén transzkripciója indul. Ez a kései gének második hulláma, mely

strukturális és a virion érésében szerepet játszó fehérjéket kódolnak.

Transzkript izoformák számszerűsítése A transzkript izoformák annotálása után abundanciájukat mértük fel. Az izoforma típusok expressziója nem egységes. Az első órában monocisztronos és splice-olt izoformákból áll a transzkriptom. A fertőzést követő második órában az izoforma diverzitás drámaian megnő. Az alternatíván termináló izoformák kivételével, melyek csak a negyedik órában jelennek meg, minden más izoforma detektálható a második órában. Elemzésünk során megtaláltuk az UL40-es transzkript splice-olt változatát, az UL40-SP1-et. Utóbbi abundanciája konstansan növekszik a fertőzés ideje alatt, míg a nem splice-olt izoforma abundanciája a fertőzés 4.-6. órájában tetőzik, miután csökkenni kezd. Érdekes, hogy a csökkenés megindulása egybeesik a splice-olt izoforma abundanciájának hirtelen emelkedésével.

Nukleotid módosulások az RNS-ben A módosított nukleotidok detektálását az AcMNPV direkt RNS adatain végeztük a Tombó szoftvert használva. A lehetséges fals-pozitív helyek kiszűrése után 319 potenciális 5^mC -t tartalmazó pozíciót kaptunk, mely a vírus transzkriptomra térképeződött. Egy lehetséges metilációs konszenzus szekvenciát is találtunk, az UUACCG-t (a módosított bázis aláhúzva), illetve arra jutottunk, hogy az 5^mC módosulások C és G-gazdag szekvencia környezetben történik. A vírus tizenkét

transzkriptjében kiterjedt metiláltságot találtunk, különös tekintettel az ORF-ekre.

Diszkusszió

Tézisemben a hosszú leolvasásokat adó szekvenálás használhatóságát tárgyalom virális transzkript izoformák megtalálására, annotálására, a vírus génexpressziójának karakterizálására és az RNS nukleotidok detektálására és térképezésére. A HSV-1 transzkriptomával végzett munkánk számos új 5' és 3' UTR izoforma, multigénes és nem kódoló transzkript annotálásához vezetett. Rámutattunk, hogy ezen izoformák hatására a virális transzkriptek között fokozott számú átfedés figyelhető meg. Feltételezzük, hogy az átfedő virális transzkriptek, a már korábban *in vivo* és *in vitro* leírt ún. transzkripció interferencia szabályozást fejtik a szomszédos gének felett (Cullen, Lomedico and Ju, 1984; Martens, Laprade and Winston, 2004; Hu *et al.*, 2007). Ugyanakkor a nem transzlálódó régiók a poszttranszlációs folyamatok cisz-típusú szabályozását tehetik lehetővé (Matoulkova *et al.*, 2012), vagy az uORF-eken keresztül akár a transzlációra is hatással lehetnek (Young and Wek, 2016; Lin *et al.*, 2019). Tézisemben hangsúlyozom a több típusú könyvtárkészítési módszer és szekvenáló platform, valamint a leolvasások szűrésének fontosságát az RT és a PCR során létrejövő műtermékek eltávolításában (Cocquet *et al.*, 2006; Balázs *et al.*, 2019).

A vírus életciklusát génexpresszióján keresztül érthetjük meg. Rámutatok a hosszú leolvasásokat adó szekvenálás új generációs szekvenálással szembeni

előnyeire a transzkript izoformák strukturális és expressziós karakterizálásában. A BoHV-1 génkifejeződését nem amplifikált cDNS könyvtár használatáva elemeztük. A génexpresszió az alfaherpeszvírusokra jellemző háromosztatúságot mutat, úgy mint a HSV-1. Rámutatunk, hogy a cisz-szabályozó szekvenciákkal rendelkező TSH és TVH-k általában jobban expresszálódnak, mint azok, melyek elől hiányzik a szabályozó szekvencia. Ugyanakkor a TATA box jelenléte növeli bármely 5' nem transzlálódó régió izoforma kifejeződését. Karakterizáltuk a circ gén expresszióját, mely a kanonikusan korai génekkel megegyező időben és mértékben fejeződik ki. Ez utalhat a gén transzkripcióban vagy replikációban betöltött szerepére.

A RNS nukleotid módosulásaival kapcsolatos tudásunk bővítése érdekében jel-analízis vizsgálatot végeztünk az AcMNPV transzkriptomából készített ONT direkt RNS könyvtáron. Gazdag citozin metiláltságot detektáltunk, különösen C és G gazdag régiókban, melyet egy korábbi kutatás emlősökben már kimutatott (Yang *et al.*, 2017). Egy lehetséges metilációs jelet is találtunk, az UUACCG-t, azonban ezt további kutatásoknak kell megerősítenie. Yang és mtsai. igazolták, hogy az ^{5m}C nukleotidok az ALYREF adapter fehérjéken keresztül fokozzák az mRNS-ek nukleáris exportját emlős sejtekben. Boyne és mtsai. ugyanezt igazolták Kaposi sarcoma-asszociált herpeszvírusban (Boyne, Colgan and Whitehouse, 2008). Az ALYREF az ízeltlábúakban is jelen van (Shi *et al.*, 2017). Feltételezzük, hogy a C-k fokozott metiláltsága a virális

transzkriptek nukleáris exportjában és poszttranszlációs változásainak irányításában játszhat szerepet.

Irodalomjegyzék

Adams, M. *et al.* (1991) 'Complementary DNA sequencing: expressed sequence tags and human genome project', *Science*, pp. 1651–1656. doi: 10.1126/science.2047873.

Adams, M. D. *et al.* (1995) 'Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence', *Nature*, 377(6547 Suppl), pp. 3–174.

Balázs, Z. *et al.* (2019) 'Template-switching artifacts resemble alternative polyadenylation', *BMC genomics*, 20(1), p. 824.

Blissard, G. W. and Rohrmann, G. F. (1990) 'Baculovirus Diversity and Molecular Biology', *Annual Review of Entomology*, pp. 127–155. doi: 10.1146/annurev.en.35.010190.001015.

Boyne, J. R., Colgan, K. J. and Whitehouse, A. (2008) 'Recruitment of the Complete hTREX Complex Is Required for Kaposi's Sarcoma–Associated Herpesvirus Intronless mRNA Nuclear Export and Virus Replication', *PLoS Pathogens*, p. e1000194. doi: 10.1371/journal.ppat.1000194.

Cocquet, J. *et al.* (2006) 'Reverse transcriptase template switching and false alternative transcripts', *Genomics*, 88(1), pp. 127–131.

Crick, F. (1970) 'Central Dogma of Molecular Biology', *Nature*, pp. 561–563. doi: 10.1038/227561a0.

Crick, F. H. (1958) 'On protein synthesis', *Symposia of the Society for Experimental Biology*, 12, pp. 138–163.

Cullen, B. R., Lomedico, P. T. and Ju, G. (1984) 'Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis', *Nature*, 307(5948), pp. 241–245.

Delhon, G. *et al.* (2003) 'Genome of bovine herpesvirus 5', *Journal of virology*, 77(19), pp. 10339–10347.

Depledge, D. P. *et al.* (2019) 'Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen', *Nature communications*, 10(1), p. 754.

Djebali, S. *et al.* (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414), pp. 101–108.

- Fraefel, C., Ackermann, M. and Schwyzer, M. (1994) 'Identification of the bovine herpesvirus 1 circ protein, a myristylated and virion-associated polypeptide which is not essential for virus replication in cell culture', *Journal of virology*, 68(12), pp. 8082–8088.
- Garrity, D. B., Chang, M.-J. and Blissard, G. W. (1997) 'Late Promoter Selection in the Baculovirusgp64 Envelope Fusion ProteinGene', *Virology*, pp. 167–181. doi: 10.1006/viro.1997.8540.
- Gokhale, N. S. *et al.* (2016) 'N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection', *Cell host & microbe*, 20(5), pp. 654–665.
- Ho, C. K., Gong, C. and Shuman, S. (2001) 'RNA triphosphatase component of the mRNA capping apparatus of Paramecium bursaria Chlorella virus 1', *Journal of virology*, 75(4), pp. 1744–1750.
- Hubbard, K. S. *et al.* (2013) 'Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs', *F1000Research*, 2, p. 35.
- Hu, X. *et al.* (2007) 'Transcriptional interference among the murine beta-like globin genes', *Blood*, 109(5), pp. 2210–2216.
- Hu, Y.-C. (2005) 'Baculovirus as a highly efficient expression vector in insect and mammalian cells', *Acta pharmacologica Sinica*, 26(4), pp. 405–416.
- Kogan, P. H., Chen, X. and Blissard, G. W. (1995) 'Overlapping TATA-dependent and TATA-independent early promoter activities in the baculovirus gp64 envelope fusion protein gene', *Journal of virology*, 69(3), pp. 1452–1461.
- Lavi, S. and Shatkin, A. J. (1975) 'Methylated simian virus 40-specific RNA from nuclei and cytoplasm of infected BSC-1 cells', *Proceedings of the National Academy of Sciences of the United States of America*, 72(6), pp. 2012–2016.
- Lichinchi, G. *et al.* (2016) 'Dynamics of the human and viral m6A RNA methylomes during HIV-1 infection of T cells', *Nature Microbiology*. doi: 10.1038/nmicrobiol.2016.11.
- Lin, Y. *et al.* (2019) 'Impacts of uORF codon identity and position on translation regulation', *Nucleic acids research*, 47(17), pp. 9358–9367.
- Liu, H. *et al.* (2019) 'Accurate detection of mA RNA modifications in native RNA sequences', *Nature communications*, 10(1), p. 4079.
- Macdonald, S. J. *et al.* (2012) 'Genome Sequence of Herpes Simplex Virus 1 Strain KOS', *Journal of Virology*, pp. 6371–6372. doi: 10.1128/jvi.00646-12.
- Martens, J. A., Laprade, L. and Winston, F. (2004) 'Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene', *Nature*, 429(6991), pp.

571–574.

- Matoulkova, E. *et al.* (2012) ‘The role of the 3’ untranslated region in post-transcriptional regulation of protein expression in mammalian cells’, *RNA Biology*, pp. 563–576. doi: 10.4161/rna.20231.
- Morris, K. V. and Mattick, J. S. (2014) ‘The rise of regulatory RNA’, *Nature Reviews Genetics*, pp. 423–437. doi: 10.1038/nrg3722.
- Moss, B. *et al.* (1977) ‘5’-Terminal and internal methylated nucleosides in herpes simplex virus type 1 mRNA’, *Journal of virology*, 23(2), pp. 234–239.
- Muylkens, B. *et al.* (2007) ‘Bovine herpesvirus 1 infection and infectious bovine rhinotracheitis’, *Veterinary Research*, pp. 181–209. doi: 10.1051/vetres:2006059.
- d’Offay, J. M., Fulton, R. W. and Eberle, R. (2013) ‘Complete genome sequence of the NVSL BoHV-1.1 Cooper reference strain’, *Archives of Virology*, pp. 1109–1113. doi: 10.1007/s00705-012-1574-6.
- Pokhriyal, M. *et al.* (2018) ‘Three newly identified Immediate Early Genes of Bovine herpesvirus 1 lack the characteristic Octamer binding motif- 1’, *Scientific reports*, 8(1), p. 11441.
- Proudfoot, N. J. (2016) ‘Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut’, *Science*, 352(6291), p. aad9926.
- Rajcáni, J., Andrea, V. and Ingeborg, R. (2004) ‘Peculiarities of herpes simplex virus (HSV) transcription: an overview’, *Virus genes*, 28(3), pp. 293–310.
- Roundtree, I. A. *et al.* (2017) ‘Dynamic RNA Modifications in Gene Expression Regulation’, *Cell*, pp. 1187–1200. doi: 10.1016/j.cell.2017.05.045.
- Schaefer, M., Kapoor, U. and Jantsch, M. F. (2017) ‘Understanding RNA modifications: the promises and technological bottlenecks of the “epitranscriptome”’, *Open Biology*, p. 170077. doi: 10.1098/rsob.170077.
- Shi, M. *et al.* (2017) ‘ALYREF mainly binds to the 5’ and the 3’ regions of the mRNA in vivo’, *Nucleic Acids Research*, pp. 9640–9653. doi: 10.1093/nar/gkx597.
- Sommer, S. *et al.* (1976) ‘The methylation of adenovirus-specific nuclear and cytoplasmic RNA’, *Nucleic acids research*, 3(3), pp. 749–765.
- Stoiber, M. *et al.* (2016) ‘De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing’. doi: 10.1101/094672.
- Tao, X. Y. *et al.* (2013) ‘The Autographa californica Multiple Nucleopolyhedrovirus ORF78 Is Essential for Budded Virus Production and General Occlusion Body Formation’, *Journal of Virology*, pp. 8441–8450. doi: 10.1128/jvi.01290-13.

- Tirumuru, N. *et al.* (2016) 'N(6)-methyladenosine of HIV-1 RNA regulates viral infection and HIV-1 Gag protein expression', *eLife*, 5. doi: 10.7554/eLife.15528.
- Wang, E. T. *et al.* (2008) 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, pp. 470–476. doi: 10.1038/nature07509.
- Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics*, pp. 57–63. doi: 10.1038/nrg2484.
- Wu, Z. *et al.* (2019) 'NormExpression: An R Package to Normalize Gene Expression Data Using Evaluated Methods', *Frontiers in Genetics*. doi: 10.3389/fgene.2019.00400.
- Yang, X. *et al.* (2017) '5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an mC reader', *Cell research*, 27(5), pp. 606–625.
- Young, M. D. *et al.* (2012) 'Differential Expression for RNA Sequencing (RNA-Seq) Data: Mapping, Summarization, Statistical Analysis, and Experimental Design', *Bioinformatics for High Throughput Sequencing*, pp. 169–190. doi: 10.1007/978-1-4614-0782-9_10.
- Young, S. K. and Wek, R. C. (2016) 'Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response', *The Journal of biological chemistry*, 291(33), pp. 16927–16935.
- Zhao, B. S., Roundtree, I. A. and He, C. (2018) 'Publisher Correction: Post-transcriptional gene regulation by mRNA modifications', *Nature Reviews Molecular Cell Biology*, pp. 808–808. doi: 10.1038/s41580-018-0075-1.