COMPREHENSIVE CHARACTERIZATION OF VIRAL TRANSCRIPTOMES USING LONG-READ SEQUENCING

PhD Thesis Booklet

Norbert Moldován

Department of Medical Biology
Doctoral School of Interdisciplinary Medicine
Faculty of Medicine
University of Szeged
Supervisor: Prof Zsolt Boldogkői, PhD, DSc

Szeged
- 2020 -

List of the publications directly related to the thesis

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.

Introduction

The phenotype of an organism is determined by the expression of its genes, which is modified by the environment. Crick outlined the central dogma of biology (Crick, 1958, 1970), explaining the flow of information from genes to proteins, the carrier being the RNA. The sum of the RNA molecules in a given time point, called the transcriptome characterizes the cell in which they were expressed. RNA-seq has become a ubiquitous tool used for analysing the quantitative changes of gene expression between experimental groups (differential gene expression or DGE) (Young *et al.*, 2012) or during longitudinal sampling of microorganisms and tissues (Hubbard *et al.*, 2013). It allows the structural characterization of the transcriptome, uncovering alternative splicing events (Wang *et al.*, 2008) and length isoforms (Depledge *et al.*, 2019). It broadened our understanding of the regulation of gene expression by non-coding RNAs (Djebali *et al.*, 2012; Morris and Mattick, 2014). Additionally, it is used to detect the increasing number of RNA modifications (Schaefer, Kapoor and Jantsch, 2017; Liu *et al.*, 2019). The first instances of RNA sequencing took place during the early times of Sanger sequencing, also known as first-generation sequencing (Adams *et al.*, 1991, 1995). This platform suffered from several technical difficulties due to sequencing length constraints and low throughput. Innovations in microfluidics and nanotechnology brought forward the next generation of sequencing platforms. These are capable of sequencing millions of

cDNA molecules at the same time (massively parallel sequencing), reducing the total time and cost of sequencing and enormously increasing the amount of output information. In contrast to first-generation platforms, the second generation is sensitive to the expression levels of splice isoforms and can be used to find novel genes and non-coding RNAs (Wang, Gerstein and Snyder, 2009). We are in the midst of the third revolution of sequencers, which are capable of producing long reads while maintaining high throughput.

Pacific Biosciences developed a new platform based on previously explored concepts, but with the capability of sequencing single molecules, resulting in unprecedented read lengths. A new competitor in the sequencing industry is Oxford Nanopore Technologies. They based their approach on totally new grounds by the use of protein nanopores embedded in a synthetic membrane and electrical current detectors.

Sequencing is preceded by RNA extraction and library preparation, both of which can introduce artefacts. Data pre-processing needs to filter these false products, while the analysis has to make sense of the huge amount of data produced by modern sequencers.

The HSV-1: The herpes simplex virus type 1 is belonging to the *Alphaherpesvirinae* subfamily of the *Herpesviridae.* It is one of the most widespread viruses in the human population causing cold sores or in severe cases acute encephalitis. It has the ability to establish life-long latency in its host. The 152 kbp long linear,

double-stranded DNA genome of the HSV-1 is enclosed into an icosahedral capsid, wrapped into a lipid envelope. The viral genome consists of a Unique Long and a Unique Short region, both flanked by inverted repeat region (IRL and IRS) (Macdonald *et al.*, 2012). The products of the 72 viral genes expressed during lytic infection were previously analysed by short-read sequencing (Rajcáni, Andrea and Ingeborg, 2004).

The BoHV-1: A non-human pathogenic relative of the HSV-1 is the *Bovine alphaherpesvirus type 1* (BoHV-1), affecting feedlot and dairy herds all over the world. The virus causes respiratory and fertility-related symptoms in cattle and is the main causative agent of infectious bovine rhinotracheitis (IBR) (Muylkens *et al.*, 2007). The BoHV-1's 135 kbp long double-stranded DNA genome has a high, 72% GC content, and is comprised of two unique sequences, the UL and the US, the latter being bracketed by inverted repeat regions (the internal INR and the terminal TRL repeats) (d'Offay, Fulton and Eberle, 2013). The 73 genes of the virus are mostly homologous to other *alphaherpesvirus'* genes and their nomenclature follows that of the HSV-1's. Two peculiar exceptions from this are the *circ* gene, which is common in *Varicelloviruses*, and the *ul0.5*, which is specific to BoHV-1 (Delhon *et al.*, 2003). The kinetics of the BoHV-1 gene expression are similar to other alphaherpesviruses, giving rise to immediate-early (IE), early (E), and late (L) genes. IE genes regulate the viral gene expression, E genes initiate and direct replication, while L genes are involved in the virion morphogenesis.

The AcMNPV: The *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) is an insect virus from the *Baculoviridae* family (Blissard and Rohrmann, 1990). It is widely used in recombinant protein expression systems as a gene delivery vector and as a biopesticide (Hu, 2005). Its 133 kbp double-stranded circular DNA genome harbours 156 tightly-spaced open reading frames. Viral gene expression is grouped into three distinct phases: early (E), late (L) and very late (VL). The promoter of early genes are homologous to the canonical TATA box and is recognised by the host's transcription machinery, with a frequent transcriptional initiation site being a C<u>A</u>GT (the +1 nt is underlined) sequence, which is common among insects (Kogan, Chen and Blissard, 1995). L and VL genes, however, are transcribed by the virus' own RNAP, and start at a very conserved T<u>A</u>AG (the +1 nt is underlined) motif also known as the late initiation sequence (LIS) (Garrity, Chang and Blissard, 1997).

Epitranscriptomics is an ever-increasing field of study. In addition to the 7-methylguanylate (m^7G) cap RNAs contain more than 60 other modified bases (Zhao, Roundtree and He, 2018), including other methylated bases, like m^6A, m^1A and m^5C, the isomerised pseudouridine (Ψ) or the oxidized forms of m^5C (hm^5C) (Roundtree *et al.*, 2017). Previous modification studies detected RNA methylation in simian virus-40 (Lavi and Shatkin, 1975), adenoviruses (Sommer *et al.*, 1976) and the human alphaherpesvirus 1 (HSV-1) (Moss *et al.*, 1977). Despite the wide range of known nucleotide

modifications, their functions are poorly understood. Several recent studies reported on the role of 6-methyladenosine in the replication and virus-host interactions of the HIV-1 (Lichinchi *et al.*, 2016; Tirumuru *et al.*, 2016), and the RNA synthesis of *Flaviviridae (Gokhale et al., 2016).* Viral RNAs are methylated by the cell's own methyltransferases present in both the cytoplasm and the nucleus (Gokhale *et al.*, 2016) as well as viral methyltransferases (Ho, Gong and Shuman, 2001; Tao *et al.*, 2013). No previous studies were focusing on the epitranscriptome of this widely-used viral agent.

Aims

In my thesis, I demonstrate the potential of long-read sequencing technologies in the analysis of viral transcriptomes, using three viruses studied by our group. Taking as an example the transcriptome of the HSV-1 I show the capability of third-generation sequencing technologies to explore the isoform diversity of viral RNAs and to compare the transcriptome structurally, the BoHV-1 will be used to indicate their suitability in the analysis of viral gene expression, while for the AcMNPV I will focus on their ability to detect RNA editing and modifications. Although these will be my points of convergents, for the sake of demonstration I will bring examples from other viruses studied by our group.

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. A fraction of HSV-1 and BoHV-1 samples underwent rRNA depletion, a fraction of HSV-1 and AcMNPV samples were cap-selected while another fraction from all three samples was poly(A) selected. PacBio IsoSeq libraries were prepared from the HSV-1 and AcMNPV samples, MinION amplified full-length cDNA libraries and direct RNA libraries were prepared for all three viruses, while a non-amplified full-length cDNA library was prepared for a longitudinal expression analysis experiment for the BoHV-1, followed by sequencing on PacBio and MinION sequencers.

The sequencing data was base called and pre-processed, followed by transcript isoform annotation using the LoRTIA software suite. Read abundances for the expression analysis of BoHV-1 genes and transcript isoforms were normalized using a modified version of the Median Ration Normalization of the DESeq2 software suite (Wu *et al.*, 2019). Nucleotide base modifications were detected using the Tombo software suite (Stoiber *et al.*, 2016).

Results

Annotating the viral transcriptome: We found a total of 182 novel putative mRNAs, 8 new ncRNAs, 53 new TSS and TES isoforms of the HSV-1 transcripts. Another potential of LRS technologies is the accurate detection of splice isoforms. Among many, we discovered new splice variants of the UL34-35 and the RL1-RL2 transcript. We identified 201 multigenic transcripts, many of which are polycistronic, while several are so-called complex transcripts overlapping antisense ORFs.

The increased number of multigenic transcripts, TSS and TES isoforms results in several novel overlaps in the HSV-1 transcriptome. These overlaps are the result of read-through events between parallelly or convergently oriented transcripts, or the long 5' UTRs of RNAs in divergent orientation. Practically all convergent genes produce overlaps caused by the RNAPII continuing its transcription for a short time following mRNA cleavage (Proudfoot, 2016). However, these residual RNA molecules are short-lived as exonucleases degrade them quickly. Another source of convergent overlaps is alternative termination and cleavage. We could observe transcripts produced by this phenomenon in low abundance in many HSV-1 genes.

Analysis of the viral gene expression: Gene expression analysis was performed on the direct cDNA sequencing data of the BoHV-1 transcriptome. The abundance of TSSs with a TATA box shows a higher increase than those lacking a TATA box, the latter plateauing after 4h

p.i. Also, TESs with a canonical PAS have a higher increase in abundance, and at 12 h p.i. their number exceeds TESs without a PAS three-fold. Gene expression for the BoHV-1 follows the general pattern specific to alphaherpesvirus. We found that the viral genes create at least three temporal classes, in accordance with the previous classification used for alphaherpesviruses: IE, E and L. Genes (bicp0, bicp22) playing an essential role in viral transcription are expressed at the first hour of the infection. We found that CIRC, the mRNA of a myristylated tegument protein with unknown function (Fraefel, Ackermann and Schwyzer, 1994) is also expressed at this time point, suggesting a role in viral transcription or replication. Pokhriyal et al. reported (Pokhriyal *et al.*, 2018) the immediate early expression of three genes (ul21, ul33 and ul34) ought to be thought late. We could detect two of these (ul33 and ul34) to exhibit an early expression pattern, and in addition, found that 41 out of the 69 analysed genes are already expressing at 2h p.i., however, 66 only reach their maximum abundance following replication, suggesting that some early genes benefit from the onset of DNA replication. Twenty-three genes, including ul12, essential for the processing of the newly synthesised DNA and many structural components of the tegument and the capsid start their expression at 4h p.i. and continue to be expressed during the infection cycle. These represent the first wave of late genes and are encompassing the most abundant viral genes. The expression of only four of the viral genes seems to be dependent on DNA replication, with an

expression start following 4h p.i. These genes could be considered a second wave of the late genes, and encode structural components of the virion and proteins playing a role in the viral egress.

Quantitative analysis of transcript isoforms Following transcript isoform annotation we evaluated their change in abundance. We observed heterogeneity in the expression of different isoform types. In the first hour of the infection, the transcriptome is composed of only monocistronic and spliced isoforms. However, starting from the second hour the isoform diversity increases dramatically, with alternatively terminating mRNAs being the only isoform type not appearing until 4h p.i. We detected a novel splice variant of UL40 the UL40-SP1, which results in a frameshift mutation and a putative protein with altered amino acid composition. The abundance of UL40-SP1 shows a constant and steady increase during the infection, while the non-spliced isoform, has a peak around 4-6h, after which it's abundance decreases. Intriguingly, this decrease coincides with the increase in abundance of it's spliced isoform.

RNA modifications 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 319 putative $5^{m}C$ sites mapped to the viral transcriptome. We detected a potential methylation consensus sequence: UUA<u>C</u>CG (the modified base underlined), and we found that in general

5^{m}Cs were abundant in C and G-rich contexts. Twelve of the viral transcripts showed extensive methylation, with most of the modified Cs found in the ORFs.

Discussion

In my thesis, I discussed the use of LRS for the discovery and annotation of viral transcript isoforms, its capability for the characterization of gene expression, and the detection of RNA modifications. Our work on the HSV-1 transcriptome yielded a number of novel 5' and 3' UTR isoforms, multigenic and non-coding RNAs. We show that as a result of these isoforms the viral transcriptome has a highly overlapping nature. We hypothesize that overlapping viral transcripts could have a regulatory effect on their neighbouring genes through transcriptional interference, similarly to which was demonstrated by others both *in vivo* and *in vitro (Cullen, Lomedico and Ju, 1984; Martens, Laprade and Winston, 2004; Hu et al., 2007)*. At the same time, the length of the UTRs can modulate post-transcriptional processing of the mRNA through cis-acting elements (Matoulkova *et al.*, 2012), or even translation by the means of uORFs (Young and Wek, 2016; Lin *et al.*, 2019). I emphasize the importance of the use of multiple library preparation techniques and platforms, and of read filtering, to eliminate possible RT or PCR artefacts (Cocquet *et al.*, 2006; Balázs *et al.*, 2019).

The viral life cycle can be understood through the analysis of viral gene expression. In my thesis, I point out the superiority of LRS over NGS for the characterization of transcript isoforms and present our

study on the quantitation of BoHV-1 mRNAs. Non-amplified cDNA libraries were used to characterize the gene expression of the virus, which resembles a similar three-phased pattern to the canonical alphaherpesvirus gene expression derived from HSV-1. We show that TSSs and TESs with a cis-regulating element are in general highly expressed, while those lacking one are generally less expressed. We also point out that the presence of a TATA box will facilitate the expression of 5' UTR isoforms, resulting in higher expression rates. We characterized the expression of the circ gene, which is expressed at equal levels as the canonical early transcripts, suggesting its potential function in the early phases of the viral life cycle.

To broaden our understanding of the RNA modification of viruses we used ONTs direct RNA sequencing coupled with signal-level analysis. We detected abundant methylation of cytosines across the transcriptome of AcMNPV, especially in C and G-rich contexts, which is in concordance with previous studies (Yang *et al.*, 2017). We also detected a potential signal for methylation, the UUACCG, however further studies are needed to validate this finding. Yang et al. demonstrated that 5^{m}Cs are facilitating mRNA export through the ALYREF adaptor in mammalian cells, while Boyne et al showed that the same gene plays a role in the export of viral mRNAs in Kaposi's sarcoma-associated herpesvirus (Boyne, Colgan and Whitehouse, 2008). ALYREF is also present in invertebrates (Shi *et al.*, 2017). We hypothesize that the extensive methylation of

Cs on the viral transcripts plays a role in their nuclear export and posttranscriptional modification.

References

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.