AN investigation into the existence of Oyster-Herpesvirus-1 (OsHV-1) microRNAs (miRNAs)

The following document outlines a project concerning an investigation into the existence of OsHV-1 encoded miRNAs.

1. Background:

OsHV-1 is the viral agent responsible for significant economic loss within the mollusc aquaculture industry across many regions, including Ireland. (Mandas and Salati, 2017). It is a member of the *Malacoherpesviridae* family, and like all herpesviruses, it's possible that it can exist as an asymptomatic latent infection, either in the absence of replication or with periodic low level replication) (Roizmann, 2001). The primary goal of this project is to explore novel ways to overcome some of the problems we encounter when trying to detect OsHV-1 during latency. As with other viruses, OsHV-1 DNA levels are present at very low levels in potentially latently infected hosts (Lionel et al., 2013). This has important implications for diagnostics, as it increases the chance of obtaining false negatives and underestimation of viral prevalence among latently infected populations. This is because there is no viral replication taking place at this stage.

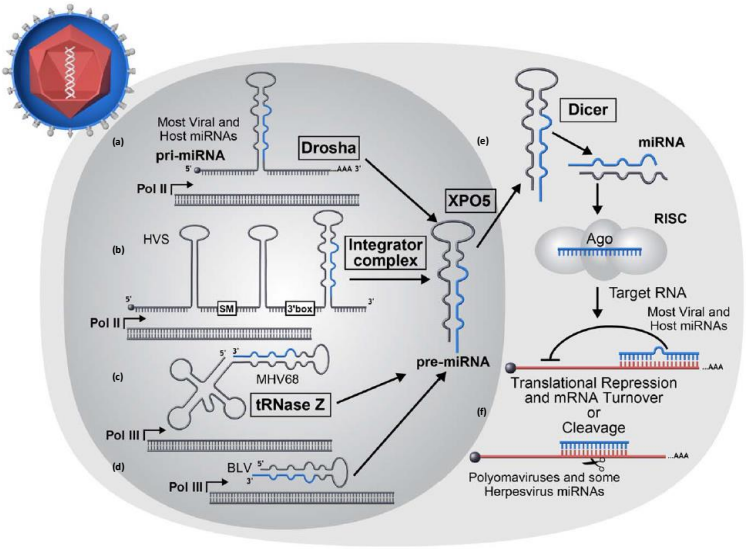
However, as observed with other herpesviruses, some OsHV-1 genes should still be transcribed during latency. Latency represents a unique transcriptional state for the virus, during which viral protein gene expression is often limited (or completely absent). Instead, some herpesviruses rely on non-protein coding genes which (unlike protein coding genes) remain highly expressed during this stage, often functioning as gene regulators involved in the long-term maintenance of latency. These are generally referred to as latency associated transcripts (Jiang et al., 2011; Lung et al., 2013; Roizmann, 2001).

Interestingly, latency associated transcripts (either non-coding or protein-coding) may be of considerable diagnostic value. This is because in theory any RNA from actively transcribed genes should be much more abundant (in terms of copy numbers) than the low levels of latent viral genome from which they are transcribed from. Thus, OsHV-1 latency associated transcripts may present us with much more appropriate diagnostic targets during latency, where viral genomic DNA levels remain low and static. While both coding and non-coding transcripts would suffice in this context, for many reasons non-coding transcripts identified during lytic infections are more likely to remain expressed during latency and furthermore they are also more likely to be present at higher levels (Fu et al., 2014; Jiang et al., 2011; Lung et al., 2013; Sullivan, 2008). For these reasons, we hope to identify OsHV-1 encoded latency associated transcripts and evaluate their diagnostic value during OsHV-1 latency.

In terms of non-coding latency associated transcripts, viral microRNAs (miRNAs) have been shown to be very prominent during latency (Grundhoff and Sullivan, 2011; Pfeffer et al., 2005; Sullivan, 2008; Sullivan and Ganem, 2005). MiRNAs are small (~21 nt) single stranded non-coding RNAs that are involved in gene regulation. Viral miRNAs can act as important regulators of host and viral gene expression during both lytic and latent infections (Cullen, 2011; Grundhoff and Sullivan, 2011; Kincaid and Sullivan, 2012; Skalsky and Cullen, 2010) and they are involved in processes such as immune evasion (Nachmani et al., 2009; Stern-Ginossar et al., 2007), host cell survival (Choy et al., 2008; Suffert et al., 2011) and lytic cycle regulation (Barth et al., 2008; Grey et al., 2007; Jurak et al., 2010).

MiRNAs are derived from longer primary miRNA transcripts (pri-miRNAs) which are processed by the nuclear RNase III enzyme, Drosha, to release precursor miRNA (pre-miRNA) hairpins (~60-70 nt) (Han et al., 2006; Lee et al., 2003). These are transported from the nucleus to the cytoplasm where they are cleaved near the terminal loop by another RNase III enzyme, Dicer, thus releasing a mature double-stranded miRNA duplex (Chendrimada et al., 2005; Ji, 2008; Koscianska et al., 2011). One of these strands is subsequently incorporated into a multi-protein complex known as the RNA-Induced

Silencing Complex (RISC) (Filipowicz et al., 2008; Khvorova et al., 2003; Schwarz et al., 2003). Once incorporated into the RISC, miRNAs base pair with specific target messenger RNAs (mRNAs), resulting in gene specific silencing either through enhanced transcript degradation or translational inhibition, mediated by RISC components (Gu and Kay, 2010; Huntzinger and Izaurralde, 2011; Su et al., 2009). This process (along with some unusual non-canonical miRNA biogenesis pathways used by some viruses, MHV68, HVS and BLV) are illustrated below.



In the diagram above, the nucleus is represented by the dark grey shape on the left. It is surrounded by the cytoplasm in light grey. (a) The canonical miRNA biogenesis pathway used by most cellular and viral miRNAs is (b) HVS pre-miRNA forming after 3' box on primary HSUR transcript and processed by cellular integrator complex (c) MHV68 pre-miRNA forming on the 3' end of a viral tRNA-like transcript (transcribed by RNA polymerase III) and processed by tRNase Z (d) BLV pre-miRNA (pri-miRNA) directly transcribed as a short discrete transcriptional unit by RNA polymerase III (e) Export of pre-miRNA from nucleus by Exprotin5 (XPO5), Dicer processing releasing mature miRNA duplex (with 3' overhangs) and incorporation into RISC (f) Targeted regulation of specific mRNAs. Top: imperfect complementary pairing to target site leading to translational repression of target mRNA, Bottom: perfect complementary base pairing to target site leading to cleavage of target mRNA. The diagram was adapted from Kincaid and Sullivan (2012).

Each miRNA is usually accompanied by several less abundant 'isomiRs' that display differing degrees of terminal heterogeneity)mainly on the 3' end) (Morin et al., 2008). In addition to miRNAs, other small non-coding RNAs known as miRNA-offset-RNAs (moRNAs) can also be processed from some pre-miRNAs (Bortoluzzi et al., 2011; Langenberger et al., 2009; Shi et al., 2009).

MiRNA genes appear to be far more prevalent among members of the *Herpesviridae* family than any other major viral group (Cullen, 2011; Grundhoff and Sullivan, 2011; Kincaid and Sullivan, 2012; Skalsky and Cullen, 2010). This may be a consequence of their distinctive biological characteristics, including large double-stranded DNA (dsDNA) genomes, nuclear replication, and their hallmark capacity to establish long-term latent infections, during which there is a need to maintain a cellular homeostasis that facilitates latency without using (or limiting the use of) potentially immunogenic viral proteins.

Although very distantly related, members of the *Alloherpesviridae and Malacoherpesviridae* families (consisting of herpesviruses of fish, reptile and mollusc hosts) possess the same biological characteristics as members the *Herpesviridae* family and thus, it is quite plausible that these viruses also encode miRNAs of their own. For these reasons, we previously investigated this possibility with *Cyprinid Herpesvirus-3* (CyHV-3), a member of the *Alloherpesviridae* family. In this case, we were able

to predict and identify at least 2 CyHV-3 encoded pre-miRNAs (and corresponding mature miRNAs) and several more putative pre-miRNAs (Donohoe et al., 2015). Given their potential as biomarkers for latent infections, in this project we are attempting to do the same with OsHV-1.

MiRNA genes are generally not conserved between different viral species. Instances of conservation or high sequence similarity have only been observed between closely related viruses (Chen et al., 2011; Jurak et al., 2010; Riley et al., 2010; Seo et al., 2008; Takane and Kanai, 2011; Walz et al., 2010) but this is not always the case (Grey et al., 2008; Walz et al., 2010; Yao et al., 2007). Thus, the prediction and identification of miRNA genes in viral genomes requires a *de novo* approach.

2. De novo prediction of miRNA genes

Since viral miRNA sequences are highly non-conserved, *de novo* prediction of miRNA genes is more useful in this context. This is a more indirect approach to miRNA gene prediction. Such strategies involve exhaustive searches of viral genomes for regions potentially giving rise to RNA secondary structures loosely resembling precursor-miRNAs (pre-miRNAs).

As genomes will produce many random RNA hairpin structures for many biological reasons and only a small number of these may function as pre-miRNAs, thus it is necessary to employ additional approaches to eliminate non-pre-miRNA-like hairpins (background) from resulting data, allowing the identification of the most likely pre-miRNA predictions. Relevant cut-off points for attributes associated with VMir pre-mRNA predictions were established previously based on the attributes of a dataset of 92 known viral pre-miRNAs (Donohoe et al., 2015). The same cut-off points were applied to pre-miRNA predictions in the OsHV-1 genome.

2.1 Approach:

Pre-miRNA prediction on the OsHV-1 genome (GenBank Ac. AY509253.2) was carried out using VMir. The genome was analysed using a 500 bp sliding analysis-window, advancing up the genome in 10 bp increments (i.e. so that each analysis-window had a 490 bp overlap with adjacent widows). All predicted RNA secondary structures occurring in each window were analysed for the presence of hairpins. Each hairpin structure was then scored on the basis of its structural consistency with pre-miRNAs. This is done by assigning values to each structural property based on a predefined scoring system, where the presence of structural properties are weighted differently depending on weather they are common features on known pre-miRNAs with penalties for non-pre-miRNA like features such as large asymmetric or symmetric bulges on the hairpin stem. VMir also used to record the amount of analysis-windows that each hairpin was detected in on the basis that those detected in a greater amount of analysis windows are more stable within their local sequence context, such stability is essential if pre-miRNAs are to form within their larger primary-miRNA transcript (which would be comprised of the local sequence surrounding the hairpin).

For this study we used Vmir (Grundhoff et al., 2006) to predict pre-miRNA-like secondary structures in the OsHV-1 genome. To distinguish between low and high probability pre-miRNA hairpins, appropriate cut-off points for scores and window-counts were established. These cut-off points were established following the analysis of six other herpesvirus genomes known to collectively encode 94 known pre-miRNAs (viruses). VMir analysis of additional these genomes, correctly predicted 92/94 of these known pre-miRNAs. Cataloguing prediction attributes associated with these 92 pre-miRNAs, gave an indication of how VMir interpreted genuine pre-miRNAs in terms of scores and window counts (Donohoe et al., 2015). This allowed the establishment of relevant cut-off points for both of these attributes which could then be applied to prediction from the OsHV-1 genome, thus improving the quality of the predictions.

2.2 Results:

VMIr analysis of the OSHV-1 genome resulted in the prediction of 25,960 hairpins (Figure 1). In order to reduce background noise within this data the following cut-off points were applied, Minimum Hairpin Score: 123, Minimum Hairpin Window Count: 39, Minimum Hairpin Length: 50. This eliminated all but 94 hairpins (Figure 2). These 94 hairpins, display a high degree of consistency with the dataset of 92 known herpesvirus pre-miRNAs analysed in a previous study (Donohoe et al, pending publication) in terms of structural features and stability, and thus out of the 25,960 predicted hairpins, these are the ones most likely to represent genuine OSHV-1 encoded pre-miRNAs.

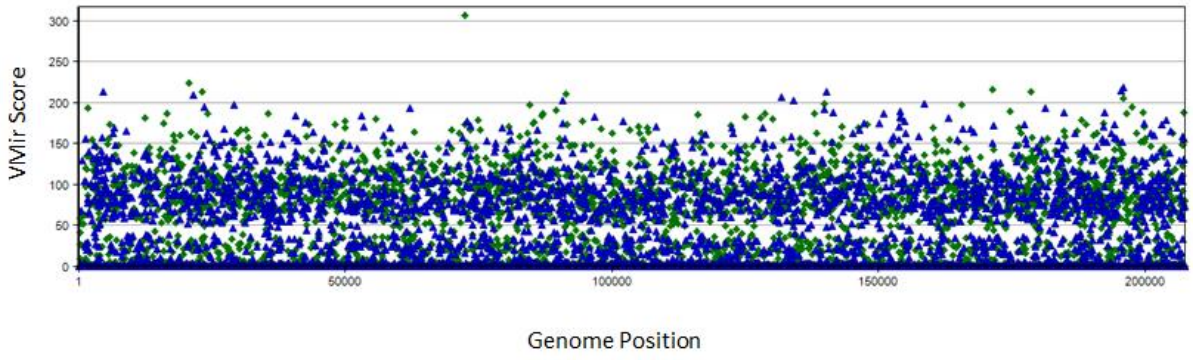


Figure 1. Output from VMir analysis of the OsHV-1 genome before application of cut-off points
A total of 25,960 hairpin structures were predicted. Hairpins on the forward and reverse strands of the OsHV-1 genome are represented by blue triangles and green diamonds respectively.

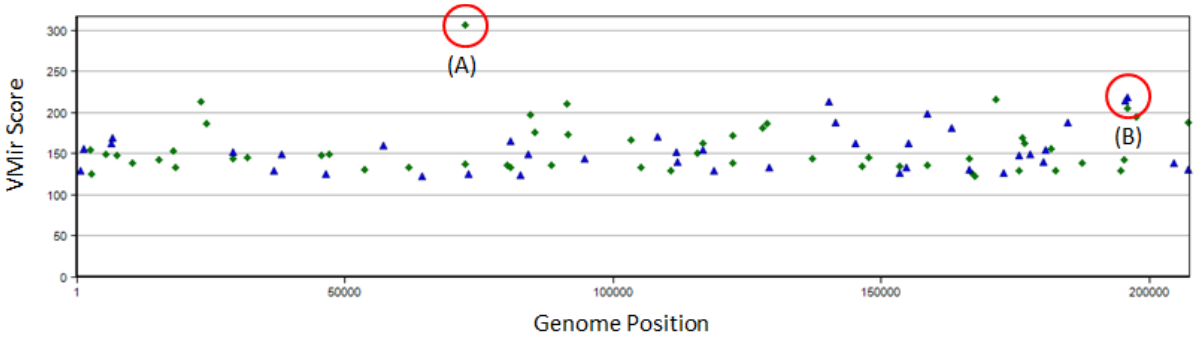
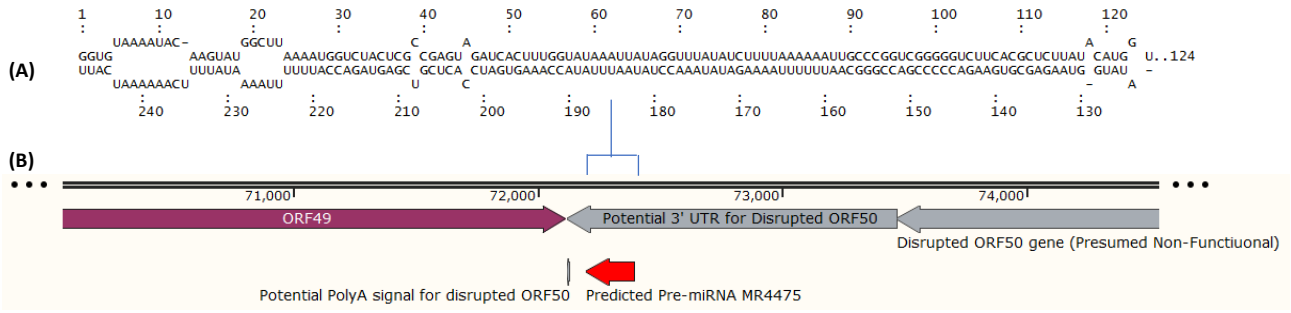


Figure 2. Output from VMir analysis of the OsHV-1 genome after application of cut-off points
A total of 94 hairpins remained after application of the cut-off points. Hairpins on the forward and reverse strands of the OsHV-1 genome are represented by blue triangles and green diamonds respectively (A) Position of MR4475 (B) Position of MD12288 and MD12304

Notably, one of these predicted hairpins, designated MR4475, (Figure 2A) had a particularly high score of 306.8. This is far higher than what has previously been observed in any other viral genome analysed in this manner (previously the highest scoring pre-miRNA observed was mdv1-mir-M3 in Mareks Disease virus, scoring 276.3) (Donohoe et al., 2015). This suggests that this predicted hairpin, is particularly pre-miRNA-like. Interestingly, the same stem loop structure has been previously identified a possible OsHV-1 origin of replication, owing to the presence of a triple palindrome, forming an almost perfect duplex stem with very few mismatches (Davison et al., 2005). It is not unprecedented for palindromic sequences associated with viral origins of replication to also function as pre-miRNAs as observed with Herpes-Simples 1 and 2 (Jurak et al., 2010). For MR4475 to act as an OsHV-1 pre-miRNA, it would first be necessary for this loci to be transcribed into RNA. Interestingly, a closer look at this loci (Figure 3) suggests that its position, downstream from the disrupted ORF50 gene. While this gene cannot be translated into protein, it may very well be still transcribed. We were able to identify a potential polyadenylation signal (marking the limit of the ORF50 3' UTR) downstream of the predicted pre-miRNA MR4475, indicating that this loci may indeed be transcribed, thus transcription of the disrupted ORF50 gene may actually give rise to a primary-miRNA transcript containing the predicted pre-miRNA MR4475. Notably this would indicate that a single transcript may once have

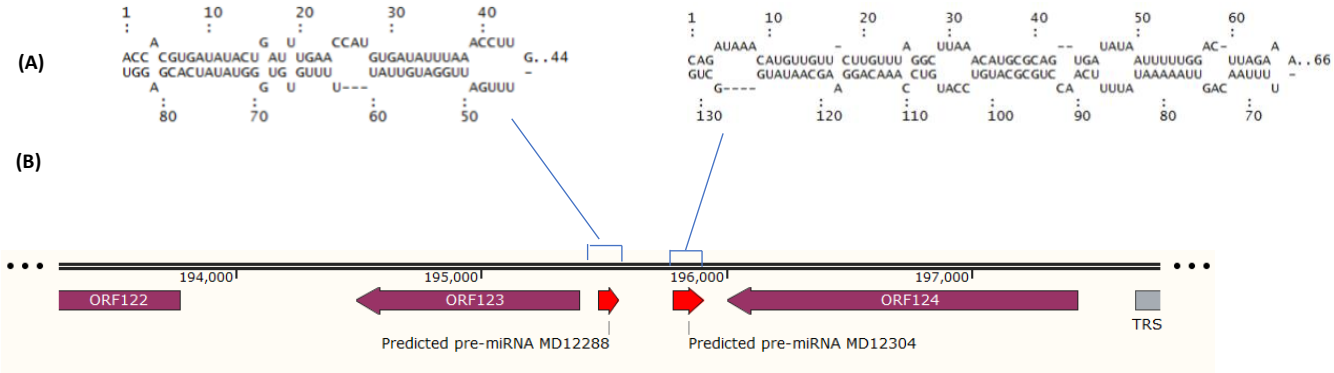
functioned as both an mRNA (for a once active ORF50 protein) and a primary-miRNA, as has been observed with other viral mRNAs (Buck et al., 2007; Cai et al., 2005).



OsHV-1 genome (GenBank Ac. AY509253.2) position 70,720 to 75199

Figure 3. Genomic context of predicted pre-miRNA MR4475
(A) Predicted structure of pre-miRNA MR4475 (B) MR4475 is located on the reverse strand of the OsHV-1 genome, downstream of ORF50 and upstream of the potential PolyA signal for this gene indicating it is within the potential 3' UTR of this mRNA. The same stem-loop structure is also predicted to be an OsHV-1 origin of replication.

Also, of note within the data are two high scoring predicted pre-miRNAs, MD12288 and MD12304, occurring in close proximity on the OsHV-1 genome (Figure 2B and Figure 4). The fact that they occur in close proximity is significant as viral miRNAs mainly tend to occur close together in clusters i.e. several pre-miRNAs are usually derived from the same primary-miRNA transcript. Interestingly, they are situated a non-coding region opposite what is likely to be the 3' UTR of ORF124 (Figure 4B). Thus, in theory miRNAs derived from these predicted pre-miRNAs would be perfectly complementary to the ORF124 3' UTR and may act to regulate the expression of this viral gene.



OsHV-1 genome (GenBank Ac. AY509253.2) position 192,104 to 199,590

Figure 4. Genomic context of predicted pre-miRNAs MD12288 and MD12304
(A) Predicted structure of pre-miRNA MD12288 and MD12304 (B) These predicted pre-miRNAs are located on the forward strand of the OsHV-1 genome, opposite what is likely to be the 3' UTR of ORF124 on the opposite strand.

To install VMir - consisting of VMIR Analyser (to predict pre-miRNAs in genomes) and VMIr Viewer (to view predictions file) the setup file can be downloaded from here. Once installed, VMir viewer can be used to open the OsHV-1 predictions. The OsHV-1 file corresponding to the pre-miRNA predictions mentioned in this document is available for download here.

In similar analysis of the CyHV-3 genome using the same approach to pre-miRNA prediction, four high-probability CyHV-3 pre-miRNAs identified through sequencing of small RNAs were also included in the top 25 CyHV-3 pre-miRNA VMir predictions (Donohoe et al., 2015), indicating the effectiveness of both

our noise reduction strategy and the VMir scoring algorithm. Thus, based on the OsHV-1 predictions, we concluded that there was a reasonable chance that this virus also encoded miRNAs. For this reason, we explored this possibility further through sequencing of small RNA from OsHV-1 infections.

3. Evaluation of OsHV-1 miRNA predictions through small RNA sequencing from OsHV-1 infected tissue

3.1 Important considerations:

The identification of novel miRNAs is carried out though deep sequencing of small RNAs. Due to high evolutionary conservation, for most eukaryotic organisms, this involves the identification of small RNA sequences that are homologous to known miRNA sequences and derived from regions of the genome that are predicted to fold into pre-miRNAs. These miRNAs, should be derived from the stem of the pre-miRNA hairpin structures, adjacent to the loop (see example of identification of miRNA form small RNAseq data in Figure 5)

However, the identification of novel *non-conserved* miRNAs requires more evidence. A piece of key evidence supporting the presence of novel non-conserved miRNAs are the presence of "miRNA-like alignment signatures" to the predicted pre-miRNAs. Crucially, in addition to the detection of the miRNA itself, the identification of such signatures requires the detection of the minor from (partner strand derived from opposite arm of pre-miRNA duplex) and various isomiRs (variants of the dominant small RNA, with heterogenous 3' ends, by-products of miRNA biogenesis) all of which will be present in much lower levels (see figure 5). For this reason, when using RNAseq to identify miRNAs, it's necessary to enrich the samples for small RNAs within the miRNA size range (i.e. 17-25nt). This RNA size fraction contains many random degradation products form non-miRNA transcripts (mRNA, rRNA etc), however miRNAs (host and viral) will dominate this size fraction. This is because unlike other small RNAs present (degradation products), miRNAs and their isomiRs are protected from degradation while incorporated into the RISC, and thus can accumulate in high levels compared to non-miRNA small RNAs within this fraction. Thus, specific enrichment within this size fraction represents a primary distinction between miRNAs and background noise caused non-miRNA small RNAs. This enrichment is typically done through PAGE size fractionation and subsequent gel purification.

In addition to the miRNAs (major and minor form) and associated isomiRs, some pre-miRNAs may show evidence of moRs, which are derived further away form the loop but adjacent to miRNAs (see Figure 5). Sometimes fragments of the loop may also be detected, but these are present in much lower quantities as unlike mature miRNAs, small RNA fragments derived form the loop they are typically not incorporated into the RISC and thus not protected from degradation, so much like other random RNA fragments, they do not accumulate in high quantities in 17-25nt fractions. If you align small RNAs representing miRNAs (and isomiRs), moRNAs and loop fragments to a pre-miRNA sequence, they will form a distinct stack-like pattern with minimum overlap between stacks. This referred to as a "five-phased-precursor" (Berezikov et al., 2011; Bortoluzzi et al., 2012, 2011; Donohoe et al., 2015) and a good example is shown in Figure 5.

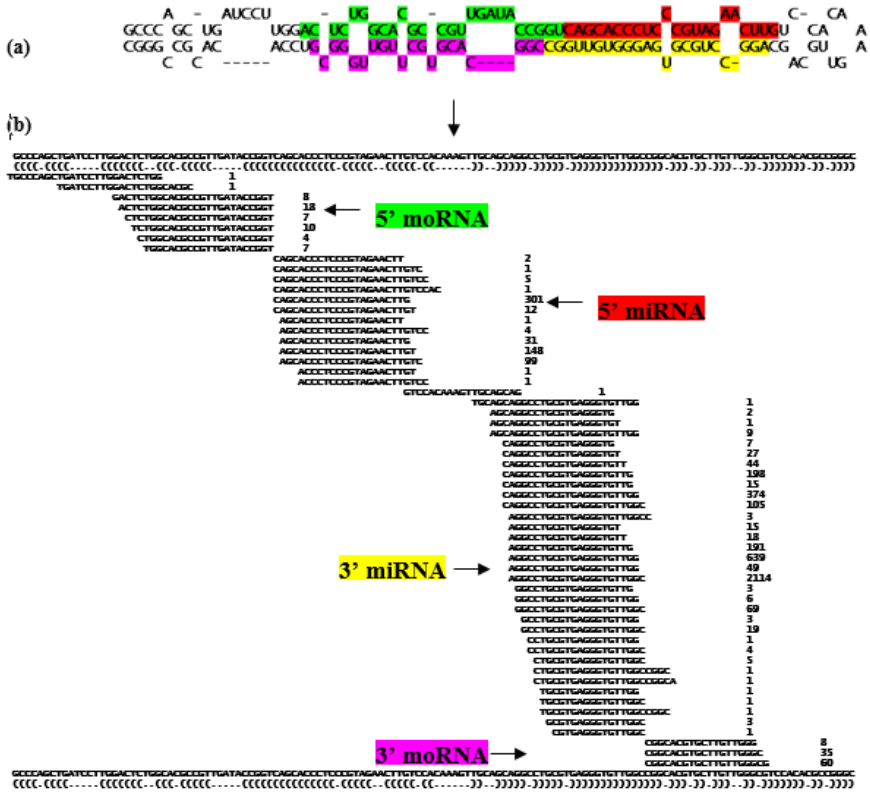


Figure 5 "Five-phased precursor" miRNA-like alignment signature displayed by CyVH-3 pre-miRNA candidate MD11776 (based on alignment signatures in Supplementary Dataset J from Donohoe *et al* 2015)
(a) Secondary structure of pre-miRNA candidate MD11776 (predicted by VMir) with miRNA and moRNA reads highlighted on the stem (b) Alignment signature of small RNAs to pre-miRNA candidate MD11776. This is a good example of a site that displays all of the different types of distinct functional groups that are expected to be derived from a single pre-miRNA (miRNAs, moRNAs, terminal loop and associated isomers). Based on the read count information, it can be seen that it is the dominant read at each locus that is designated as the miRNA/moRNA with the rest designated as isomiRs/ moRNA isomers. The collective alignment of miRNAs, IsomiRs and moRs, should form distinct "stacks" representing each small RNA derivative, with minimum overlap between stacks.

Such patterns indicate that the pre-miRNAs were processed in a manner consistent with miRNA biogenesis - for more information see the following publications (Donohoe et al., 2015; Kozomara and Griffiths-Jones, 2011). By contrast, small RNA reads that represent random degradation products from larger transcripts are more likely to map to genomes (or falsely predicted pre-miRNAs) in the form of overlapping reads offset across the entire length of the degraded transcript due to the fact that their degradation is an entirely random process, rather than an ordered RNAse III cleavage events based on structural motifs, see illustration in Figure 6

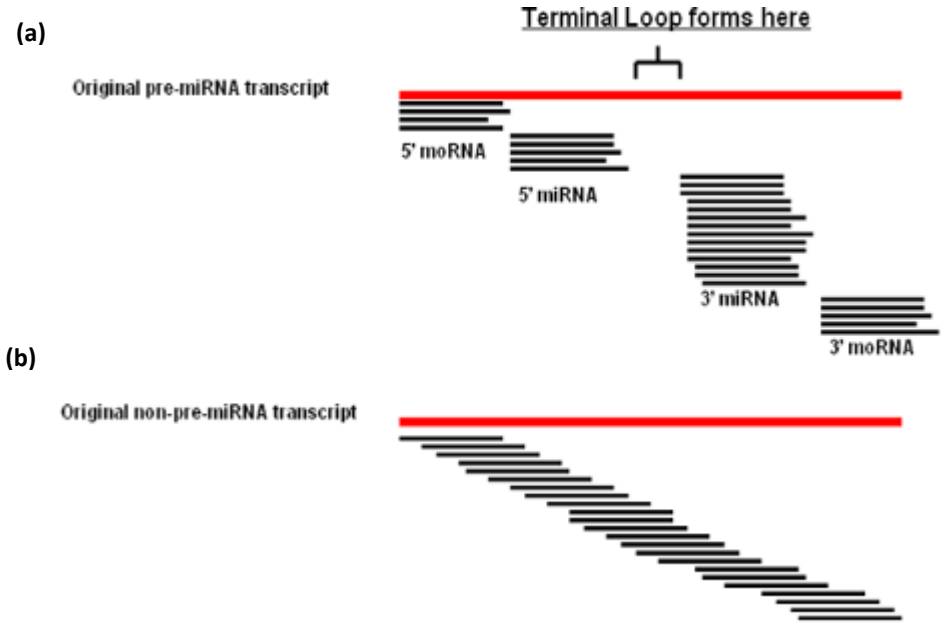


Figure 6. Examples of miRNA-like and non-miRNA-like alignment signatures
(a) miRNA-like alignment signature, miRNAs, isomiiRs form "stacks" directly adjacent to the terminal loop, moRs (if present) for stacks adjacent to these (b) non-miRNA-like alignment signature, overlapping reads offset across the entire length of the degraded transcript due to the fact that their degradation is an entirely random process, rather than an ordered RNAse III cleavage based on structural motifs.

The examination of alignment signatures is one of the easiest ways of distinguishing between non-miRNA and miRNA sequences in small RNA sequencing datasets, thus it is very useful in the identification of novel non-conserved miRNAs. Programs such as MiRDeep (Friedländer et al., 2012, 2008) MIREAP (sourceforge.net, 2013) have been developed to identify these kinds of small RNA alignment signatures in RNAseq data and proven to be very useful in the identification of CyHV-3 encoded miRNAs (Donohoe et al., 2015)

When sequencing RNA from virally infected cells *in vitro* and *in vivo*, typically host RNA reads will dominate the data but this is much more extreme in in vivo, where with viral reads being much lower compared to *in vitro* infections (Amen and Griffiths, 2011; Umbach et al., 2009), presumably due to enhanced viral replication in vitro in the absence of major elements of the immune response.

As viral reads will be intrinsically low, there is a chance that if genuine novel miRNAs are present in the data, the lower abundance minor form and isomiRs (necessary to support annotation of novel miRNAs) corresponding to these miRNAs will not be sequenced due to insufficient depth. Therefore, when choosing samples, it is necessary select those with the highest amount of viral RNA relative to host RNA.

3.2 Samples used in this study

The samples used for small RNAseq consisted of gill tissue from pacific Oysters (*Crassostrea gigas*). Thirty gill samples were selected on the basis of testing positive for OsHV-1 by PCR. We then extracted total RNA from these samples using Trizol. All total RNA samples screened to identify those with the highest amount of viral RNA relative to host RNA. To do this, we used RT-PCR to measure the levels of OsHV-1 transcript ORF87 and normalised them to levels of Oyster EF-1 alpha. This acted as an estimation of viral RNA relative to host RNA in each sample. A single sample referred to as "Sample 1" was identified as having the highest amount of viral RNA and thus selected for small RNAseq analysis.

3.3 RNA sample Preparation
17-25nt RNA was isolated from Sample 1 total RNA. This was done as per Donohoe *et al* (2015). The presence of miRNAs in the 17-25nt sample was confirmed by detecting known host miRNAs using stem-loop PCR. This sample was then sent to EMBL for sequencing.

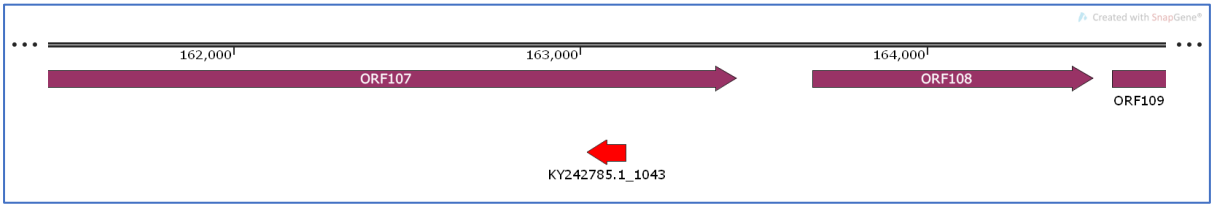
3.3 RNA sequencing
Illumina small RNA sequencing libraries were prepared and applied to a single lane on an Illumina NextSeq Unit, yielding 441,291,224 small RNA reads (Single End, 75bp length, strand-specific, sense orientation). This data was retrieved in standard fastq format.

3.4 Pre-processing of sequencing data
Fastqc analysis of the fastq file showed that there was a large amount of adapter sequence present in the data, presumably due to the short RNA insert length (~17-25nt) and large read length (Fastqc output available here) These were removed using bbduk (form BBMap package). Fastqc analysis after revealed no remaining adaptor sequence. The settings and code used are here along with details of additional small RNA sequencing adaptors that were added to the reference file (Fastqc output post bbduk trimming is available here)

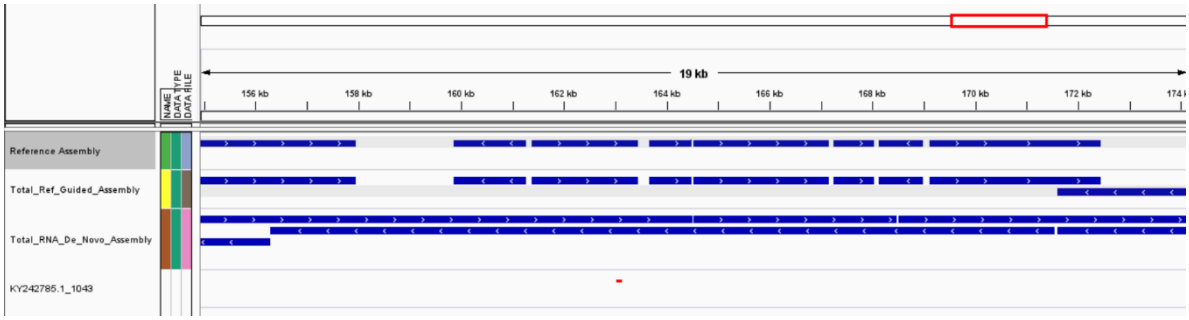
3.5 Results of all analysis so far:
miRdeep2 was used are per guidelines. Briefly, Bowtie was used to index the viral genome (Genbank Ref: KY242785.1) This genome index was then used as a reference for the mapper.pl module of the miRDeep2 package. As part of pre-processing prior to mapping, all reads <17nt in length were removed, the fastq was parsed into .fasta format and the file was then collapsed into a fasta file containing unique reads only with read counts information (i.e. the amount of times it appears in the fastq file) incorporated into the fasta identifier in the format x_<read count>. The output files form the mapping process (collapsed reads and .arf ile with mapping information) were the utilised with the miRDeep2 core algorithm. All code used is available here. All input and output files available here in this folder (its best to download this entire folder)

MiRDeep2 examined the genomic regions that small RANs were mapping to and checked if these regions may give rise to pre-miRNA-like hairpin structures. It then checked for miRNA-like alignment signatures to these hairpin structures. It identified many hairpins (structures and reads available in output.mrd.xlsx available here) however it only two of these had mirDeep2 scores >0. These are available in this html file "result_27_03_2019_t_09_14_46.html", with links to pdfs of their structures. These two pre-miRNAs are referred to as "KY242785.1_1043" and "KY242785.1_54" They are described below. The locations of these were viewed in IGV along with the reference assembly. In order to potentially sequence associated primary RNA transcripts and thus provide evidence to support the annotation of novel RNAs we also sequenced total RNA form the same infection. This was mapped and reassembled using Hisat2 and StringTie (Pertea et al., 2016). The transcriptome was reassembled using reference-guided and de novo methods. Both were used to look for evidence of novel long non-coding RNAs overlapping "KY242785.1_1043" and "KY242785.1_54" putative pre-miRNAs. These assemblies can be viewed in IGV by loading the appropriate GTFs (GTFs from Reference Guided and De Novo assembly are available here and here respectively). The reference can also viewed as a separate track (GTF derived from Genbank GFF which is here).

Only one of these putative pre-miRNAs, KY242785.1_1043, falls within a non-coding region. This putative pre-miRNA identified by miRDeep2 is located at position 163055-163111 on the reverse strand of the genome opposite to ORF107 (non-coding part of the region) see below.



As we can see in from the *de novo* assembly there is evidence of a large non-coding transcript overlapping this region on the same strand. However as with all de novo assemblies caution should be taken in interpreting this a non-coding transcript without some experimental evidence (RT PCR etc), as this is important in order to rule it out as an artefact of the assembly process, see below.



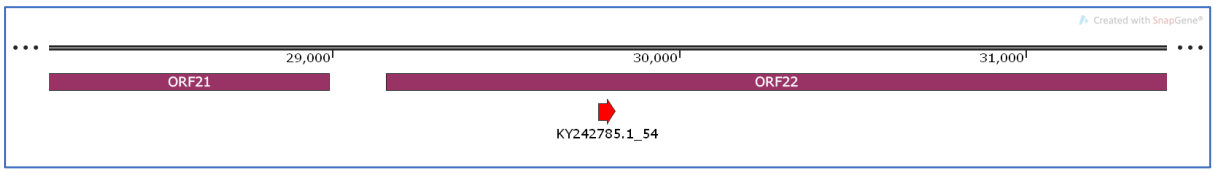
The same pre-miRNA (although longer) was also predicted in VMir analysis (MR10366, score 143.3), see structure below (also from VMir analysis)

```
1               10          20          30          40          50
:               :           :           :           :           :
    CA-   ------        G   CAU   UC   CC       -     UC       UCAAUUG
GGCG    GGU       GCAACAG UU    AU   CA   UCCAUG CCGG   UGUGGA         U..56
CUGC    CCA       CGUUGUU AA    UG   GU   AGGUAC GGUC   ACAUCU         -
    ACA   UUCUGG        G   UC-   GA   UA       U     CU        UAUACAU
        :           :           :           :           :           :
        110         100         90          80          70          60
```

However, if we examine the alignment signature in the output.mrd.xlsx file we can see that there is only one small RNA aligned to it with a read count of 16. As there is no minor form mapping to the other arm of the pre-miRNA and no isomiRs, this provides little evidence that this is actually a genuine miRNA.

```
pri_seq          gcaacagguucauauuccaccuccaugccggucuguggaucaauuguuacauauucuacauccuggucauggaauugaggucuaaguuguugcggucuuaccacacgucuc #NM
pri_struct       ((((((((.((....(((....((((((((((..((((((.....................))))))..))))).))))))...)))....))).))))))(((...))).......... 
seq_257143245_x16 ....................cuccaugccggucuguggauc.................................................................... 0
```

The other putative miRNA, KY242785.1_54, comes from a coding region, occurring within ORF22



The same pre-miRNA (although longer) was also predicted in VMir analysis (MD1968, score: 87.7), see structure below (also from VMir analysis), but its less pre-miRNA-like with a large bulge and thus has a lower score

```
:           :           :           :
      U   CAUAAC---   A   G     U         U
GCCAUU GC           GG  AA  AAGG  GGAAGAU   A..35
UGGUAA CG           CC  UU  UUCC  UCUUCUG   -
      -   UUAACCUAA   G   G     U         C
 :           :           :           :
 70          60          50          40
```

In terms of alignment profile. While there are more unique reads mapped to this hairpin, all except one have mismatches and are thus of lower confidence. Again in the absence of minor-forms or extensive isomiRs, there is little evidence to support the annotation of this region as pre-miRNA coding.

```
exp                ffffffffffffffffffffMMMMMMMMMMMMMMMMMMMMMllllllllSSSSSSSSSSSSSSSSSSSSSfffffffffffffffffffffffffffffffffffffffff
pri_seq            cuucgagaaggccauuugccauaacggaaagaagguggaagauuacgucuucuuccuuguugccaauccaauugcaaugguauccuuagcaagaaagggauucaugagcaa
pri_struct         .........(((.....))).....((.((.((((.(((((((...))))))).)))).)).)).......((((.((((.((((((........))))))))))..)))) #MM
seq_270224795_x1   ....................auaacggaaagaaggCggaag...................................................................... 1
seq_255974649_x17  ....................auaacggaaagaagguggaag...................................................................... 0
seq_268712254_x1   ....................auaacUgaaagaagguggaag...................................................................... 1
seq_266886232_x2   ....................auaacggaaagaagguggaagaA.................................................................... 1
```

These are the only two putative pre-miRNAs to have miRDeep scores greater than zero and were thus the only ones selected by the program. However, they simply don't have enough reads mapping to them to give any evidence that they are genuine pre-miRNAs. Despite the fact that over 400million reads were sequenced from infected tissue, the viral content was still quite low, thus creating difficulties in terms of sufficient coverage to identify and existing viral miRNAs and lower abundance minor-forms and isomiRs to via pre-miRNA alignment signatures and support novel miRNA annotation

3.6 Suggestions of what else could be tried

Based on past experience, other programs and approaches to novel viral miRNA identification can sometimes uncover thigs that miRDeep will miss. Thus other things that may be considered are based on this.

Manual inspection of pre-miRNA alignment signatures
There are many more putative pre-miRNAs and alignment signatures in the same miRDeep results file. Most of them have many more reads mapping, they may be worth examining further by even manually looking for those with the best pre-miRNA like alignment signatures (as also done by Donohoe et al 2015).

Mireap or other programs
The same small RNAseq data could also be analysed using other approaches such as Mireap etc.

VMIr Repeat for KY242785.1
The previous VMir analysis was done on a different strain (genbank AY509253.2). It may be worth doing the analysis again on the microVar strain (genbank KY242785.1) as its more relevant to this infection. It may be useful to map the small RNAs to sequences of the top 100-200 intergenic pre-miRNA predictions. Alternatively (as per manual method in Donohoe et al 2015). It may be useful to take the most abundant small RNAs mapping to non-coding regions of the viral genome (more likely to be miRNAs than any other small RNAs) and use the search facility on VMir viewer to search for any hits to predicted pre-miRNAs

References

Amen, M.A., Griffiths, A., 2011. Identification and Expression Analysis of Herpes B Virus-Encoded Small RNAs. J. Virol. 85, 7296–7311. https://doi.org/10.1128/JVI.00505-11

Barth, S., Pfuhl, T., Mamiani, A., Ehses, C., Roemer, K., Kremmer, E., Jaker, C., Hock, J., Meister, G., Grasser, F.A., 2008. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. Nucleic Acids Res. 36, 666–675. https://doi.org/10.1093/nar/gkm1080

Berezikov, E., Robine, N., Samsonova, A., Westholm, J.O., Naqvi, A., Hung, J.-H., Okamura, K., Dai, Q., Bortolamiol-Becet, D., Martin, R., Zhao, Y., Zamore, P.D., Hannon, G.J., Marra, M.A., Weng, Z., Perrimon, N., Lai, E.C., 2011. Deep annotation of Drosophila melanogaster microRNAs

yields insights into their processing, modification, and emergence. Genome Res. 21, 203–215. https://doi.org/10.1101/gr.116657.110

Bortoluzzi, S., Biasiolo, M., Bisognin, A., 2011. MicroRNA-offset RNAs (moRNAs): by-product spectators or functional players? Trends Mol. Med. 17, 473–474. https://doi.org/10.1016/j.molmed.2011.05.005

Bortoluzzi, S., Bisognin, A., Biasiolo, M., Guglielmelli, P., Biamonte, F., Norfo, R., Manfredini, R., Vannucchi, A.M., 2012. Characterization and discovery of novel miRNAs and moRNAs in JAK2V617F-mutated SET2 cells. Blood 119, e120–130. https://doi.org/10.1182/blood-2011-07-368001

Buck, A.H., Santoyo-Lopez, J., Robertson, K.A., Kumar, D.S., Reczko, M., Ghazal, P., 2007. Discrete Clusters of Virus-Encoded MicroRNAs Are Associated with Complementary Strands of the Genome and the 7.2-Kilobase Stable Intron in Murine Cytomegalovirus. J. Virol. 81, 13761–13770. https://doi.org/10.1128/JVI.01290-07

Cai, X., Lu, S., Zhang, Z., Gonzalez, C.M., Damania, B., Cullen, B.R., 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc. Natl. Acad. Sci. U. S. A. 102, 5570–5575. https://doi.org/10.1073/pnas.0408192102

Chen, C.J., Kincaid, R.P., Seo, G.J., Bennett, M.D., Sullivan, C.S., 2011. Insights into Polyomaviridae microRNA function derived from study of the bandicoot papillomatosis carcinomatosis viruses. J. Virol. 85, 4487–4500. https://doi.org/10.1128/JVI.02557-10

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., Shiekhattar, R., 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436, 740–744. https://doi.org/10.1038/nature03868

Choy, E.Y.-W., Siu, K.-L., Kok, K.-H., Lung, R.W.-M., Tsang, C.M., To, K.-F., Kwong, D.L.-W., Tsao, S.W., Jin, D.-Y., 2008. An Epstein-Barr virus-encoded microRNA targets PUMA to promote host cell survival. J. Exp. Med. 205, 2551–2560. https://doi.org/10.1084/jem.20072581

Cullen, B.R., 2011. Viruses and microRNAs: RISCy interactions with serious consequences. Genes Dev. 25, 1881–1894. https://doi.org/10.1101/gad.17352611

Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Deuff, R.-M.L., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J. Gen. Virol. 86, 41–53. https://doi.org/10.1099/vir.0.80382-0

Donohoe, O.H., Henshilwood, K., Way, K., Hakimjavadi, R., Stone, D.M., Walls, D., 2015. Identification and Characterization of Cyprinid Herpesvirus-3 (CyHV-3) Encoded MicroRNAs. PLOS ONE 10, e0125434. https://doi.org/10.1371/journal.pone.0125434

Filipowicz, W., Bhattacharyya, S.N., Sonenberg, N., 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9, 102–114. https://doi.org/10.1038/nrg2290

Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N., 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26, 407–415. https://doi.org/10.1038/nbt1394

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52. https://doi.org/10.1093/nar/gkr688

Fu, M., Gao, Y., Zhou, Q., Zhang, Q., Peng, Y., Tian, K., Wang, J., Zheng, X., 2014. Human cytomegalovirus latent infection alters the expression of cellular and viral microRNA. Gene 536, 272–278. https://doi.org/10.1016/j.gene.2013.12.012

Grey, F., Hook, L., Nelson, J., 2008. The functions of herpesvirus-encoded microRNAs. Med. Microbiol. Immunol. (Berl.) 197, 261–267. https://doi.org/10.1007/s00430-007-0070-1

Grey, F., Meyers, H., White, E.A., Spector, D.H., Nelson, J., 2007. A Human Cytomegalovirus-Encoded microRNA Regulates Expression of Multiple Viral Genes Involved in Replication. PLoS Pathog. 3. https://doi.org/10.1371/journal.ppat.0030163

Grundhoff, A., Sullivan, C.S., 2011. Virus-encoded microRNAs. Virology 411, 325–343. https://doi.org/10.1016/j.virol.2011.01.002

Grundhoff, A., Sullivan, C.S., Ganem, D., 2006. A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. RNA N. Y. N 12, 733–750. https://doi.org/10.1261/rna.2326106

Gu, S., Kay, M.A., 2010. How do miRNAs mediate translational repression? Silence 1, 11. https://doi.org/10.1186/1758-907X-1-11

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., Kim, V.N., 2006. Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. Cell 125, 887–901. https://doi.org/10.1016/j.cell.2006.03.043

Huntzinger, E., Izaurralde, E., 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. 12, 99–110. https://doi.org/10.1038/nrg2936

Jiang, X., Alami Chentoufi, A., Hsiang, C., Carpenter, D., Osorio, N., BenMohamed, L., Fraser, N.W., Jones, C., Wechsler, S.L., 2011. The Herpes Simplex Virus Type 1 Latency-Associated Transcript Can Protect Neuron-Derived C1300 and Neuro2A Cells from Granzyme B-Induced Apoptosis and CD8 T-Cell Killing. J. Virol. 85, 2325–2332. https://doi.org/10.1128/JVI.01791-10

Ji, X., 2008. The mechanism of RNase III action: how dicer dices. Curr. Top. Microbiol. Immunol. 320, 99–116.

Jurak, I., Kramer, M.F., Mellor, J.C., van Lint, A.L., Roth, F.P., Knipe, D.M., Coen, D.M., 2010. Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2. J. Virol. 84, 4659–4672. https://doi.org/10.1128/JVI.02725-09

Khvorova, A., Reynolds, A., Jayasena, S.D., 2003. Functional siRNAs and miRNAs Exhibit Strand Bias. Cell 115, 209–216. https://doi.org/10.1016/S0092-8674(03)00801-8

Kincaid, R.P., Sullivan, C.S., 2012. Virus-Encoded microRNAs: An Overview and a Look to the Future. PLoS Pathog 8, e1003018. https://doi.org/10.1371/journal.ppat.1003018

Koscianska, E., Starega-Roslan, J., Krzyzosiak, W.J., 2011. The Role of Dicer Protein Partners in the Processing of MicroRNA Precursors. PLoS ONE 6, e28548. https://doi.org/10.1371/journal.pone.0028548

Kozomara, A., Griffiths-Jones, S., 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152–157. https://doi.org/10.1093/nar/gkq1027

Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., Stadler, P.F., 2009. Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 25, 2298–2301. https://doi.org/10.1093/bioinformatics/btp419

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415–419. https://doi.org/10.1038/nature01957

Lionel, D., Guyader, T., Tourbiez, D., Pépin, J.-F., 2013. Is horizontal transmission of the Ostreid herpesvirus OsHV-1 in Crassostrea gigas affected by unselected or selected survival status in adults to juveniles? Aquaculture 408–409, 51–57. https://doi.org/10.1016/j.aquaculture.2013.05.025

Lung, R.W.-M., Tong, J.H.-M., To, K.-F., 2013. Emerging Roles of Small Epstein-Barr Virus Derived Non-Coding RNAs in Epithelial Malignancy. Int. J. Mol. Sci. 14, 17378–17409. https://doi.org/10.3390/ijms140917378

Mandas, D., Salati, F., 2017. Ostreid herpesvirus: A pathogen of oysters. Virol. Res. Rev. 1. https://doi.org/10.15761/VRR.1000112

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., Marra, M.A., 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 18, 610–621. https://doi.org/10.1101/gr.7179508

Nachmani, D., Stern-Ginossar, N., Sarid, R., Mandelboim, O., 2009. Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. Cell Host Microbe 5, 376–385. https://doi.org/10.1016/j.chom.2009.03.003

Pertea, M., Kim, D., Pertea, G., Leek, J.T., Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. Nat. Protoc. 11, 1650–1667. https://doi.org/10.1038/nprot.2016.095

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grässer, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., Tuschl, T., 2005. Identification of microRNAs of the herpesvirus family. Nat. Methods 2, 269–276. https://doi.org/10.1038/nmeth746

Riley, K.J.-L., Rabinowitz, G.S., Steitz, J.A., 2010. Comprehensive analysis of Rhesus lymphocryptovirus microRNA expression. J. Virol. 84, 5148–5157. https://doi.org/10.1128/JVI.00110-10

Roizmann, B., 2001. The family Herpesviridae: A brief introduction, in: Fields Virology. pp. 2381–2397.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D., 2003. Asymmetry in the Assembly of the RNAi Enzyme Complex. Cell 115, 199–208. https://doi.org/10.1016/S0092-8674(03)00759-1

Seo, G.J., Fink, L.H.L., O'Hara, B., Atwood, W.J., Sullivan, C.S., 2008. Evolutionarily Conserved Function of a Viral MicroRNA. J. Virol. 82, 9823–9828. https://doi.org/10.1128/JVI.01144-08

Shi, W., Hendrix, D., Levine, M., Haley, B., 2009. A distinct class of small RNAs arises from pre-miRNA–proximal regions in a simple chordate. Nat. Struct. Mol. Biol. 16, 183–189. https://doi.org/10.1038/nsmb.1536

Skalsky, R.L., Cullen, B.R., 2010. Viruses, microRNAs, and Host Interactions. Annu. Rev. Microbiol. 64, 123–141. https://doi.org/10.1146/annurev.micro.112408.134243

sourceforge.net, 2013. MicroRNA Discovery By Deep Sequencing [WWW Document]. URL http://sourceforge.net/projects/mireap/ (accessed 4.1.13).

Stern-Ginossar, N., Elefant, N., Zimmermann, A., Wolf, D.G., Saleh, N., Biton, M., Horwitz, E., Prokocimer, Z., Prichard, M., Hahn, G., Goldman-Wohl, D., Greenfield, C., Yagel, S., Hengel, H., Altuvia, Y., Margalit, H., Mandelboim, O., 2007. Host immune system gene targeting by a viral miRNA. Science 317, 376–381. https://doi.org/10.1126/science.1140956

Suffert, G., Malterer, G., Hausser, J., Viiliäinen, J., Fender, A., Contrant, M., Ivacevic, T., Benes, V., Gros, F., Voinnet, O., Zavolan, M., Ojala, P.M., Haas, J.G., Pfeffer, S., 2011. Kaposi's sarcoma herpesvirus microRNAs target caspase 3 and regulate apoptosis. PLoS Pathog. 7, e1002405. https://doi.org/10.1371/journal.ppat.1002405

Su, H., Trombly, M.I., Chen, J., Wang, X., 2009. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. Genes Dev. 23, 304–317. https://doi.org/10.1101/gad.1749809

Sullivan, C.S., 2008. New roles for large and small viral RNAs in evading host defences. Nat. Rev. 9, 503–7.

Sullivan, C.S., Ganem, D., 2005. MicroRNAs and viral infection. Mol. Cell 20, 3–7. https://doi.org/10.1016/j.molcel.2005.09.012

Takane, K., Kanai, A., 2011. Vertebrate virus-encoded microRNAs and their sequence conservation. Jpn. J. Infect. Dis. 64, 357–366.

Umbach, J.L., Wang, K., Tang, S., Krause, P.R., Mont, E.K., Cohen, J.I., Cullen, B.R., 2009. Identification of viral microRNAs expressed in human sacral ganglia latently infected with herpes simplex virus 2. J. Virol. 84, 1189–92.

Walz, N., Christalla, T., Tessmer, U., Grundhoff, A., 2010. A global analysis of evolutionary conservation among known and predicted gammaherpesvirus microRNAs. J. Virol. 84, 716–728. https://doi.org/10.1128/JVI.01302-09

Yao, Y., Zhao, Y., Xu, H., Smith, L.P., Lawrie, C.H., Sewer, A., Zavolan, M., Nair, V., 2007. Marek's Disease Virus Type 2 (MDV-2)-Encoded MicroRNAs Show No Sequence Conservation with Those Encoded by MDV-1. J. Virol. 81, 7164–7170. https://doi.org/10.1128/JVI.00112-07