RACE-Seq identifies the Argonaute-2 cleavage products of RNA interference-based oligonucleotides

Usher, L.

RACE-Seq identifies the Argonaute-2 cleavage products of RNA interference-based oligonucleotides

Louise Usher

A thesis submitted in partial fulfilment of the requirements of the
University of Westminster
for the degree of Doctor of Philosophy

Faculty of Science and Technology
University of Westminster

June 2018

ABSTRACT

The recent announcement of the first successful Phase III clinical trial of a RNA interference (RNAi)-based therapeutic is a major achievement in the field. Synthetic RNAi therapeutic oligonucleotides are either first cleaved by Dicer or incorporate directly into the Argonaute-2 RNA-induced silencing complex (AGO2-RISC) and directs the protein complex to homologous RNA. Cleavage of target RNA occurs opposite bases 10-11 when counting from the 5’ end of the hybridized siRNA guide strand. The capture and identification of these cleaved products by 5’ Rapid Amplification of cDNA Ends and Sanger sequencing remains the gold standard for confirming Argonaute-2 mediated RNAi cleavage.

Next Generation Sequencing of 5’ RACE has brought new insights into the biological activity of RNAi-based oligonucleotides. This work currently represents the largest undertaking using RACE-Seq to investigate AGO2-RISC-mediated activity. RACE-Seq reported the expected RISC-cleaved product for each of the oligonucleotides investigated. Additionally, RACE-Seq analysis revealed that some of the oligonucleotides could be processed into multiple active siRNA molecules. Analysis of the activity of a Dicer substrate siRNA targeting transthyretin revealed that this molecule by-passed Dicer processing but still induced RNAi activity. In examining RACE-Seq peak profiles, an on-target mechanism of action (MOA) for up to four active siRNA derived from siRNA19 is proposed. The shRNA19 RACE-Seq assay predicted that this hairpin molecule probably exists as two distinct forms, one with a 7-nucleotide loop and the other with a 5-nucleotide loop.

The project also focused on optimising the library preparation, data filtering and data presentation for RACE-Seq. A simplified, low computation data analysis pipeline was designed and used to align the filtered dataset to a reference sequence and to count the 5’ ends. RACE-Seq is presented as a suitable solution for investigating, discriminating and quantifying specific RNA cleavage events and visualizing evidence for an on-target MOA of RNAi based oligonucleotide therapeutics.

ACKNOWLEDGEMENTS

This PhD has been a truly fantastic experience and it would not have been possible without the support and guidance that I have received from many people.

I would like to thank my supervisory team. To Dr John Murphy, for taking the helm and getting me through the final year of my PhD. Your continuous support, help and advice has been invaluable. To Dr Anatoliy Markiv, for your insightful nuggets which steered me on a straighter path.

Thank you to Dr Sterghios Moschos, for the opportunity to undertake this PhD project, for your support and care. I am also grateful to Dr Nadége Presneau for her support, encouragement and helpful discussions.

A very special thank you to Dr Pamela Greenwell, for her tireless support and encouragement, and for the many helpful discussions and advise in molecular biology and qPCR, and for sharing her wisdom and time. I would also like to thank Karima Brimah for her friendship, continued support and assistance both pre-PhD and over the last four years.

I gratefully acknowledge the funding received for my PhD from University of Westminster PhD scholarship programme.

To my PhD cohort and office colleagues, both past and present, thank you for being on this journey with me. Thanks for the cake and other treats, the cheery hellos, fist-bumps and chats in the corridors, you really have been an immense support. To Jolene, Faye, Emma, Kavit, Brad, Artun, and Nasrin thanks for being there.

To my two other musketeers, Hima and Moyin, thank you for your continued support and encouragement, for the hot chocolate breaks, the late-night dinners and the just checking-on-you-texts. I couldn't have stayed sane without you.

To my nephews, Corban and Lazarus, thank you for putting up with my continued absence. To my dear friends Jane and Quinton, and to Ravelle, Kewan and Cayden, thank you for being my second family and for holding me in your heart and for feeding me!

And finally, to my mother. Thank you for being such a fantastic inspiration and for your unconditional love and support.

Declaration

I declare that all the material presented in this thesis, is wholly my own work, unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Louise Usher

PUBLICATIONS

Moschos, Sterghios A., Louise Usher, and Mark A. Lindsay. "Clinical potential of oligonucleotide-based therapeutics in the respiratory system." *Pharmacology & therapeutics* 169 (2017): 83-103.

Theotokis, Pantazis I., Louise Usher, Christopher K. Kortschak, Ed Schwalbe, and Sterghios A. Moschos. "Profiling the Mismatch Tolerance of Argonaute 2 through Deep Sequencing of Sliced Polymorphic Viral RNAs." *Molecular Therapy-Nucleic Acids* 9 (2017): 22-33.

ACKNOWLEDGEMENT OF CONTRIBUTION

I would like to acknowledge the contributions made to my PhD by Chris Kortschak and Pantazis Theotokis. Chris coded and developed the first RACE-Seq analysis programme that I initially used to analyse my data. Pantazis coded and built the final pipeline (RACE-SEQ-lite) that I eventually used to analyze my data. Thank you both for putting up with my repeated queries on using the pipelines. I'd also like to thank Damjan Temelkovski for installing the Linux packages.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3'	3 prime
5'	5 prime
AAV8	adeno-associated virus serotype
AGO2	Argonaute2
ATP	Adenosine Tri-phosphate
bp	base pair
cDNA	complementary DNA
DMEM	Dulbecco's Modified Eagles Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DNase I	Deoxyribonuclease I
dNTP	deoxyribonucleotide triphosphate
DsiRNA	Dicer substrate small interfering RNA
dsRNA	double-stranded RNA
EMCV	encephalomycarditis virus
Exo I	Exonuclease 1
FBS	Foetal Bovine Serum
FDA	The Food and Drug Administration
fLuc	firefly luciferase gene
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
H+	hydrogen ion
HCV	Hepatic C virus

IRES	internal ribosome entry site
ISP	Ion Sphere Particle
M-MuLV	Moloney Murine leukemia Virus
miRNA	microRNA
MRE	microRNA recognition elements
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
Na	Sodium
Neo	Neomycin
NGS	Next Generation Sequencing
nM	nanomolar
NS	Non-structural
NSO	non-specific oligonucleotide
nt	nucleotide
OH-	hydroxide ion
OT2	OneTouch2
P-bodies	Processing Bodies
PARE	Parallel analysis of RNA ends
PAZ	Piwi/Argonaute/Zwille domain
PKC	Protein kinase C mRNA
pM	picomolar
pre-miRNA	precursor microRNA (60-70 bp hairpin structures)
pri-miRNA	primary microRNA; clustered hairpin structures transcribed from intronic regions
Q-RT-PCR	Quantitative real time PCR

RACE	Rapid Amplification of cDNA Ends
RACE-Seq	Rapid Amplification of cDNA Ends and Next Generation Sequencing
RISC	RNA-induced silencing complex
RACE	RNA ligase mediated Rapid Amplification of cDNA Ends
RNAi	RNA interference
RNase	Ribonuclease
rSAP	Shrimp Alkaline Phosphatase
rt-qPCR	reverse transcription quantitative polymerase chain reaction
shRNA	short hairpin RNA
siRNA	small interfering RNA
SSIV III	SuperScript III
TBE buffer	Tris/Borate/EDTA buffer
TTR	Transthyretin
UTR	untranslated region
µl	microlitre
µM	micromolar

1 INTRODUCTION

1.1 RNA interference: Biology and mechanism of action

Almost 20 years has passed since Fire and Mello published their seminal paper confirming double-stranded RNA (dsRNA) as a trigger for specific gene silencing. By injecting combinations of both sense and anti-sense RNA fragments of the *unc-22* gene into *Caenorhabditis elegans*, they concluded that dsRNA fragments resulted in more potent interference compared to when sense or antisense RNA strands were injected separately. By applying the method to numerous genes and observing phenotype alteration, they confirmed that interference was directed to a specific target. Further, they linked dsRNA-mediated interference to a reduction of target mRNA and showed both systemic and germline activity (Fire et al., 1998). This work dismissed a previous observation that single-stranded RNA resulted in gene silencing in *C. elegans* (Guo et al., 1995). The work led to significant insights of dsRNAs as triggers for post-transcriptional gene silencing in plants (Hamilton and Baulcombe, 1999; Chuang and Meyerowitz, 2000; Schweizer et al., 2000). RNAi was soon identified as an innate gene silencing mechanism of plants, insects, worms, fungi and eukaryotes and soon evolved as the tool of choice for studying gene functions. With its potential as a novel therapeutic strategy, it led to the burgeoning new field of RNA therapeutics.

In mammals, suppression of gene expression via small dsRNAs (19-22 bp), termed microRNAs (miRNAs), is well characterised (Figure 1.1). The sequences encoding miRNAs are located in diverse regions throughout the genome, with nearly half of all identified miRNAs originating within introns of protein coding genes (Lee et al., 2004; Kim et al., 2012). These miRNA precursor sequences occur singly or in clusters of up to 6 miRNA precursors, although some larger clusters have also been identified. The clusters may be arranged as homo-seed clusters (miRNAs with identical ‘seed’ sequences), hetero-seed clusters (miRNA with distinct ‘seed’ sequences) or a mix of both (Wang et al., 2016). The miRNA precursors are transcribed primarily by RNA polymerase II (Lee et al., 2004; Cai et al., 2004; Saini et al., 2007) which is the catalytic complex responsible for transcription of DNA to mRNA, but miRNA precursors have also been identified within regions of RNA polymerase III transcription activity (Borchert et al., 2006; Gu et al., 2009). Thus, biogenesis of miRNA begins in the nucleus with the transcription of long, primary miRNA (pri-miRNA) that fold back in single or clustered hairpin structures with imperfect pairing (Lee et al., 2004). The top part of the hairpin structures are then excised from the pri-miRNA by the enzyme Drosha, together with its co-factor DGCR8 (also known as Pasha) to produce the 60-70 nucleotide precursor miRNAs (pre-miRNAs) (Zeng and Cullen, 2003; Cai et al., 2004) having distinct 2 nt, 3' end overhangs (Lund and Dahlberg, 2006; Helvik et al., 2007). This processing of intronic miRNAs occurs in unison with mRNA exon processing (Melamed et al., 2013; Ramalingam et al., 2014). In cases where miRNA location overlaps an exon region, miRNA processing competes with mRNA splicing factors (Melamed et al., 2013).

Exportin-5 next transports the pre-miRNAs to the cytoplasm to be further processed by Dicer, an RNase III enzyme. Human Dicer recognises the 5' end and/or the 2-nt 3' overhang of pre-miRNAs or dsRNA and cleaves ~22 nt into the stem generating mature miRNAs (Park et al., 2011; Feng et al., 2012). In humans, miRNAs may then then be loaded to any one of the four members of the Argonaute (AGO) protein family in association with other proteins, forming the RNA-induced silencing complex (RISC). The passenger strand of the duplex miRNA is released and degraded, inducing an active RISC that is guided to target mRNA by sequence homology between the miRNA guide strand and the target mRNA (Elbashir *et al.*, 2001; Cai *et al.*, 2009). In humans, miRNAs associate with AGO1, AGO2, AGO3 and AGO4. While AGO2 is able to cleave target RNA, AGO1, AGO3 and AGO4 do not have this endonuclease catalytic activity. MiRNAs generally form imperfect pairing between the guide strand and target mRNA and result in a RISC that blocks the translational machinery, thus driving gene silencing. The minimal requirement for sequence recognition between a miRNA and mRNA target is defined as the 'seed' and is the 2-8 nucleotides counted from the 5' end of the miRNA guide strand (Lewis et al., 2005).

‘Seed’ sequences, as well as their targets which contain the microRNA recognition elements (MRE) are conserved across species. In animals, extensive complementarity between target mRNA and miRNA is rare, thus a single miRNA may target hundreds of RNA transcripts. On the other hand, RISC-mediated endonucleolytic target cleavage requires extensive complementarity between the guide stand and target and is only mediated by AGO2. Endogenous derived small interfering RNA (endo-siRNA) with perfect pairing to mRNA targets are synthesised from double-stranded endogenous RNAs and incorporate into AGO2 RISC and result in target cleavage (Stein et al., 2015). Additionally, exogenously introduced dsRNA designed to have high complementarity to an mRNA target can enter the RNAi pathway at either the Dicer stage (where it is processed to siRNA) or can associate directly with AGO2, generating active RISC and results in cleavage of its target (Tuschl et al., 1999).

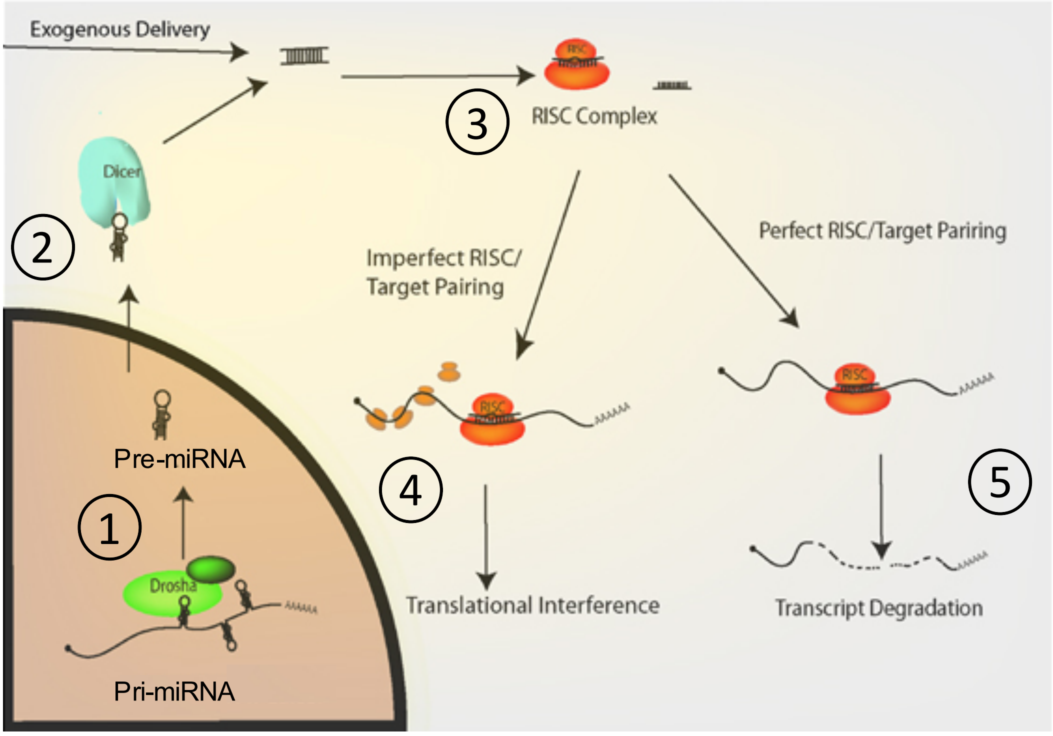


Figure 1.1 MicroRNA processing and RNA interference in mammals. Primary miRNAs (pri-miRNA) are transcribed in the nucleus and cleaved to short 60-70 nucleotide primary miRNA hairpins (1), transported to the cytoplasm where they are recognised by Dicer (2) and processed to 19-22 bp RNA duplex with characteristic 2-nucleotide 3' overhangs. Dicer processed miRNA or synthetic siRNA incorporate into the RNA induced silencing complex (RISC) where the passenger strand is released and the guide strand directs RISC to target RNA resulting in either blocking of the translation machinery (4) or Argonaute 2-mediated cleavage of target RNA (5). (Adapted from Heidersbach et al., 2006)

1.1.1 Dicer: structure and function implications for siRNA

Dicer is a critical component of the RNAi pathway and is a large (220-kDa), multi-domain, RNase III enzyme. The domains of human Dicer include; the Piwi/Argonaute/Zwille (PAZ) domain, a dsRNA binding domain, the helicase domain (so named for homology rather than function) and two RNase III-like domains (Figure 1.2 A and B). The PAZ domain recognises the characteristic 2-nt overhang on the 3' end of RNAi intermediates (Park et al., 2011; Liu et al., 2015), docking and securing the duplex RNA. The stem region of the dsRNA molecule is then metered out by the distance between the PAZ domain and the RNase III domain, which measures 65 angstrom, the length of 25 base pairs of RNA (MacRae, 2006) (Figure 1.2 C). The dsRNA binding domain and platform region also contribute to substrate recognition. The RNase III-like domains form an intramolecular dimer to create a single active catalytic site responsible for cleavage of RNA (Lau et al., 2012; Ma et al., 2012). Further evidence suggests that internal bulge/loop structures and the terminal loop of hairpin RNAs may be sensed differently by the helicase domain (Soifer et al., 2008; Gu et al., 2012). The mechanisms for Dicer recognition and processing of substrates has been intensely investigated with cell-based experiments identifying that Dicer processes plasmid expressed shRNAs into different biologically active siRNAs, (Flores-Jasso et al., 2009; Gu et al., 2012; Snead et al., 2013; Denise et al., 2014) which has the potential to affect their functional efficacy and increases the potential for off-target effects. For example, using northern blotting, Gu et al. (2012) showed that in Pol III transcription of sh-miR30 from a plasmid vector, containing a passenger strand at the 5' arm, a 9-nt loop and a 3' guide strand arm, multiple siRNA products of varying length for both the passenger strand and the guide strand were processed from Dicer cleavage. Additional deep sequencing of these small RNAs produced over 750 000 reads that mapped to the siRNA guide strand, and showed the guide strand to have two distinct start positions. One of the start positions aligned to the expected Dicer cleavage site for the shRNA, but approximately 18% of the reads showed a start position 2-nt upstream of the expected site. The authors went on to show that such non-canonical cleavages might be avoided by placing the cleavage site 2-nt from a bulge or loop in the stem of the shRNA, for more precise Dicer processing. Thus, defining the precise rules for Dicer cleavage is an ongoing effort, but is critically linked to the design of effective shRNA. Generally, both siRNA and shRNA are designed such that one strand, the guide strand, has preferential loading to RISC. However, since Dicer performs non-canonical cleavages generating siRNA duplexes with different ends this can (i) result in a shift in 'seed' region and (ii) allow for the possibility that the passenger strand may show increased preference for loading to RISC. In either case, this would increase the potential for off target effects.

1.1.2 Human Argonaute 2: Structure and function

Argonaute proteins are highly specialised small-RNA-binding modules comprising four distinct domains, namely; the N-terminus, PAZ, Mid and the PIWI domains. Structural studies have revealed that human AGO2 proteins are bi-lobed, having the N-terminus and PAZ domains as one lobe and the MID-PIWI domains as the other lobe, connected by a hinge. This hinge allows for structural re-arrangement during RISC binding (Song et al., 2004; Schirle and MacRae, 2012). The PAZ domain binds the 3' end of small duplex RNA, while the MID domain recognises the 5' phosphate at position 1 of the siRNA guide strand (the thermodynamically unstable end of the siRNA), anchoring the small RNA to the protein. The N-terminus assists in unwinding of duplex RNA, while the PIWI domain (which is structurally similar to RNase H) is involved in cleavage of the passenger strand (Matranga et al., 2005; Rand et al., 2005a) which is released generating the active RISC. The 'seed' region of the guide is arranged into the AGO2 cleft with bases 2-4 exposed to enable scanning of target mRNAs (Nakanishi, 2016) (Figure 1.3). Target recognition enables hybridization of the 'seed' and additional hybridization of the next 5-8 nucleotides beyond the 'seed' to form a helix of favourable conformation for cleavage between mRNA nucleotides 10 and 11when counting from the 5' end of the siRNA guide strand (Elbashir Martinez et al., 2001; Wang et al., 2009; Salomon et al., 2015).

In designing potent RNAi drugs, increasingly, sugar modifications and phosphorothioate backbone modification are employed to improve the pharmacokinetic properties of therapeutic siRNA. To shed light on how these modifications might impact siRNA interactions with AGO2, Schirle et al., (2006) used crystal structure analysis to compare the interactions of modified vs unmodified siRNAs with AGO2. In observing the 5' end of siRNAs, they found closely matched conformation of nucleotide positions 2-4 for modified vs unmodified siRNAs, but observed a substantial deviation in conformation of modified siRNA at position 5-8. This difference in observed conformational arrangement did not impair knockdown potency (Schirle et al., 2016). In this study, the central region could not be modelled with confidence, so the impact of conformation alteration on alignment of the siRNA with the catalytic centre, and therefore the impact on RISC cleavage precision, could not be determined. While *in vitro* biochemical studies may aid in designing potent modified siRNAs (Schlegel et al., 2017), there remains a lack of tools to specifically characterize the impact of such modifications on RISC behaviour in a biological context. The impact of chemical modification, conjugates, siRNA design and delivery strategies on target cleavage precision remains an avenue to be investigated.

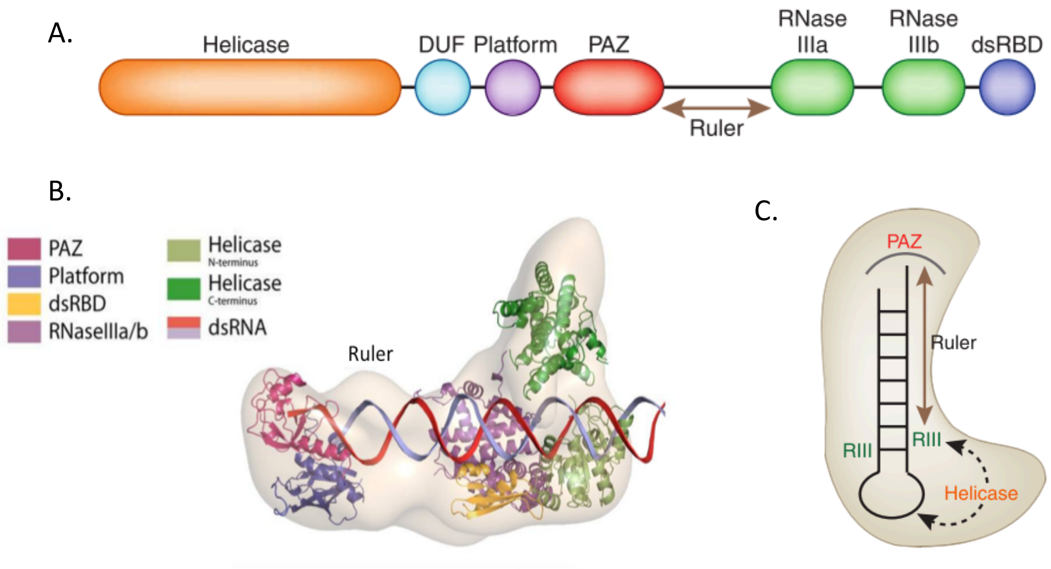


Figure 1.2 Elucidation of Dicer architecture.

Figure (A) illustrates the various domains of human dicer, with (B) illustrating the positioning of the various domains and illustrates the regions of siRNA interaction. Figure (C) illustrates how the "ruler domain" meters out a distance between the PAZ domain and RNase III cleavage domain. (Adapted from: Ipsaro and Joshua-Tor, (2015); Sawh and Duchaine, (2012))

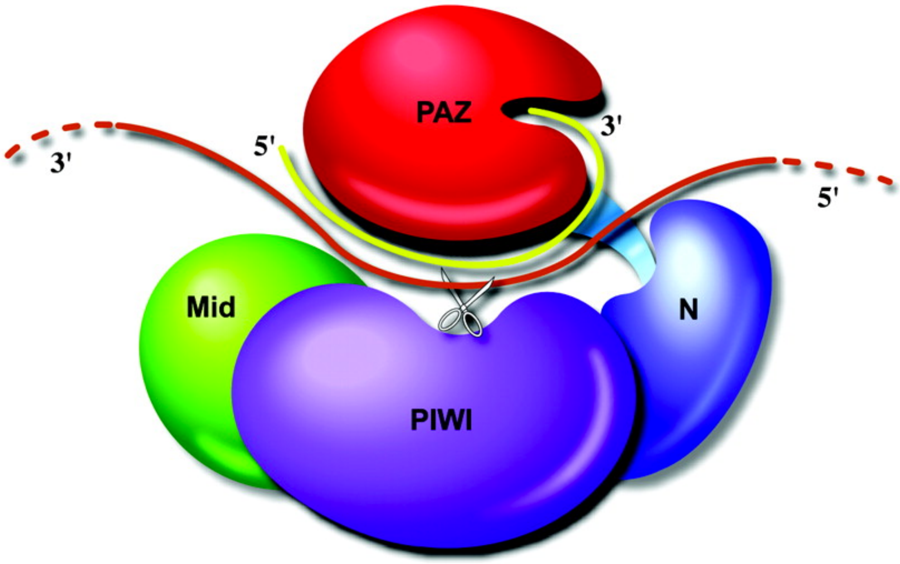


Figure 1.3 Schematic of siRNA guide strand hybridization and target cleavage.

The 3’ 2 nt overhang of the guide strand is recognised by a hydrophobic pocket of the PAZ domain. The’ seed’ region (nucleotides 2-8) are positioned for hybridization to the target. Further hybridization at 10-11 correctly positions the target mRNA for cleavage by the PIWI domain. (Yellow = siRNA guide strand, brown = mRNA target strand). (Song et al., 2004)

1.2 Targeting disease: RNAi therapeutics in the clinic

The first Phase I clinical trial of siRNA therapeutics began in 2004, only six years after the discovery of RNAi. Considerable advances in research, technological improvements and novel applications have led to numerous clinical trials and pre-clinical investigations (Tabernero et al., 2013; Schultheis et al., 2014; Patel et al., 2016; Ganesh et al., 2016). Since its first discovery and early adoption, the high expectations that these RNAi-based drugs were expected to deliver has thus far failed to achieve Food and Drug Administration (FDA) approval. To date, only six oligonucleotide therapeutics (reviewed in Stein and Castanotto, 2017) have been approved for marketing by the FDA, with no RNAi-based therapeutics currently on this list. Pre-2008, RNAi therapeutic development had focused on commercialization of intellectual property and proof of principal publications. The field took a hit when it was realised that certain RNAi formulations led to stimulation of the innate immune system (Sledz et al., 2003; Sioud, 2007; Robbins et al., 2009). Post 2011, the field of RNAi therapeutics entered a recovery phase, buoyed primarily by an increase in quality of science; steered by heightened scientific scrutiny, a drive towards delivery technologies, considerable technical improvements, a focus at reducing adverse risks in pre-clinical and clinical assessments and advances in enabling technologies.

In late September 2017, Alnylam Pharmaceuticals (in collaboration with Sanofi Genzyme) announced the first ever positive Phase III results for an RNAi therapeutic. Patients receiving their RNAi drug, Patisiran, showed significant improvement in health and drug efficacy scores compared to patients receiving a placebo, after 18 months. This announcement marks a potentially significant milestone in the journey of RNAi therapeutics to the clinic (Alnylam Pharmaceuticals press release on 20 September 2017 http://investors.alnylam.com/press-releases?mobile=1&items_per_page=20&page=2).

1.2.1 Hepatitis Virus C: Disease and molecular biology

Chronic infection by Hepatitis Virus C (HCV) results in mild disease, or can lead to severe symptoms such as scar-tissue formation and cirrhosis. Cirrhosis can lead to hepatocellular carcinoma and can ultimately result in death. Humans are the only known natural host of HCV with an estimated 3% of the worlds' population chronically infected, placing significant pressure on public health systems. HCV is classified into at least seven major genotypes, whose genomes differ by at least 30%, with genotype prevalence showing varied worldwide distribution. Genotypes are further classified into subtypes (a, b, c, d, *etc.*). Genomic characterisation of HCV is further complicated as in infected individuals, HCV circulates as a population of diverse but closely related variants referred to as "quasispecies" (Smith et al., 2014).

HCV is a single strand, positive sense, RNA virus of the family Flaviviridae, genus *Hepacivirus*. The RNA genome is of approximately 9,600 nucleotides long and encodes a single open reading frame (ORF) flanked by a 5' untranslated region (UTR) and 3’ UTR (Dubuisson and Cosset, 2014). The 5’ UTR has high secondary structure and holds an internal ribosome entry site (IRES) that facilitates RNA translation. The 5’ UTR also holds replication signals including a binding site for the liver specific microRNA, miR-122 (Jopling et al., 2005). The 3’ UTR also has secondary structures essential for virus replication. The single polyprotein transcribed from the ORF is cleaved by host and virus proteases into the structural protein (core, envelope glycoproteins 1 and 2), an ion channel protein (p7) and the seven non-structural (NS) proteins (NS2, NS3A, NS3B, NS4A, NS4B, NS5A, NS5B). NS5B is a RNA dependent RNA polymerase (RDRP) and is responsible for synthesising the complementary negative-strand RNA intermediate used as template in virus replication (Lohmann et al., 1999). This RDRP lacks proof-reading activity resulting in a high rate of error-prone replication generating quasispecies RNA.

1.2.2 HCV as a target and modulator of RNAi.

HCV has evolved to utilize the human RNAi pathway in a unique manner. Instead of being downregulated, HCV sequesters a liver microRNA, microRNA-122 (miR-122) as a replication factor. MiR-122-induced RISC cleavage enhances virus replication and translation. Two critical miR-122 sites are found in the 5’ UTR region, where hybridization of both the 5’ ‘seed’ region as well as the 3’ end of miR-122 at the two sites is required to elicit a replication advantage (Machlin et al., 2011). MiR-122-induced RISC interaction confers a conformational change to parts of the 5’ UTR, stabilizing the HCV RNA and enhancing replication (Pang et al., 2012). Further evidence for additional miR-122 interaction at various sites on the HCV genome other than the 5’ UTR was confirmed using HITS-CLIP (Luna et al., 2015).

HCV infection has been shown to sequester host factors from Processing (P)-bodies (cellular granules used for storage of cellular mRNA). Along with mRNA transcripts, P-bodies house various RNA degradation enzymes, (e.g. decapping proteins, 5’-3’ exonucleases and decapping activators), other non-sense mediated decay proteins and RNAi machinery components (AGO2, miRNA and GW182) (Parker and Seth, 2009). HCV diverts P-body components, including Dcp2, Xrn1, AGO2 and miR-122 to lipid droplets, the sites of HCV replication and assembly (Biegel and Pager, 2016).

Interestingly, a number of these components of P-bodies are well known in their roles in mRNA decay, but have been shown to be required for efficient HCV replication (Shimakami et al., 2012; Li et al., 2015). Although lipid droplet formation has been found to play central roles in both viral replication and viral assembly, the exact nature of the interaction of mammalian viruses and the host RNAi system remains unclear. HCV dependence on miR-122 has even been exploited in an anti-viral therapeutic approach that sequesters miR-122 using a locked nucleic acid-modified, DNA phosphorthioate antisense oligonucleotide (Janssen et al., 2013). Numerous studies have shown that HCV RNA can be targeted via the RNAi pathway by supplying designed siRNA or Dicer substrate RNA (Seo et al., 2003; Sagan et al., 2010; Chang et al., 2010).

1.2.3 TT-034, a DNA-directed RNAi therapeutic

TT-034 (Benitec Biopharma, Sydney, Australia) is a novel, DNA directed RNAi therapeutic against HCV. It consists of a recombinant adeno-associated virus serotype 8 (AAV8) vector that expresses three shRNAs, that target three distinct regions of the HCV genome (Figure 1.4). The expressed shRNA pro-drugs are substrates for Dicer processing, resulting in three separate antiviral-active siRNA drugs: siRNA6, siRNA19 and siRNA22. The AAV8 vector is replication incompetent, the TT-034 genome is released and is converted to a non-integrating stabilized episomal transcriptionally active double- stranded DNA in cells, allowing long-term expression of the RNAi pro-drugs. Strong evidence of pre-clinical safety, (Suhy et al., 2012) *in vitro* characterisation (Lavender et al., 2012) and in-depth evidence of confirmed mode of action obtained from 5'RACE studies (Denise et al., 2014) motivated the FDA to authorise initiation of clinical trials which began in January 2014. The clinical trials reached Phase I/IIa stage with the third patient dosed on 7 January 2015 and a further announcement on the 7 April 2015 stating that results from liver biopsies confirmed that the trial was proceeding as expected. However, by the end of February 2016, the company had announced termination of the programme due to reduced competitive advantage as more effective HCV therapeutics began to enter the market (Benitec Biopharma, Announcement https://blt.irmau.com/irm/PDF/1656_0/ UPDATEONHCV CLINICALTRIAL).

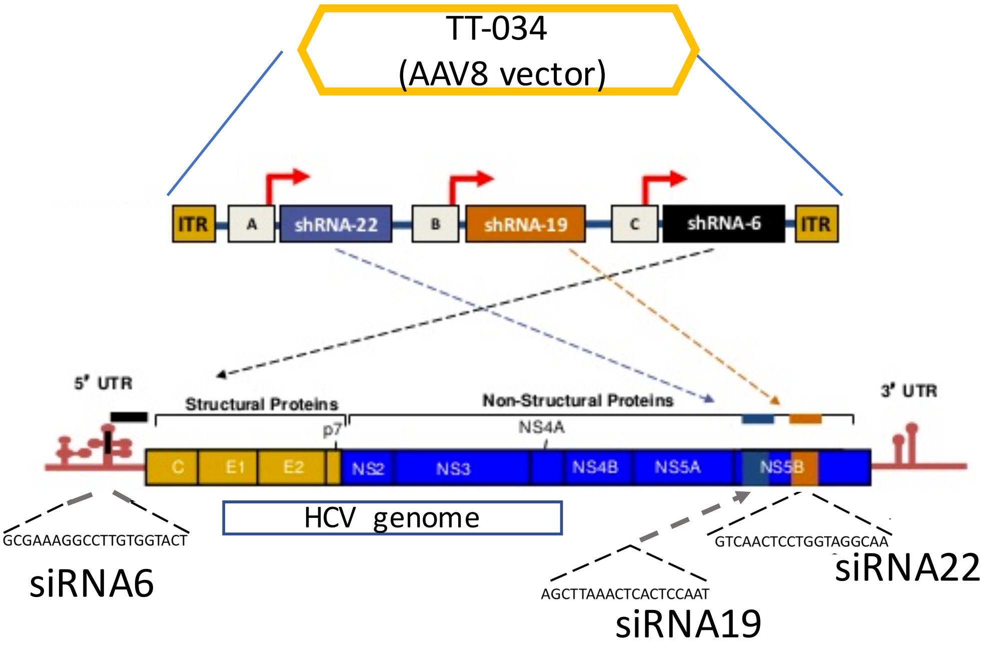


Figure 1.4 Diagrammatic representation of a DNA-directed RNAi-based drug TT-034 that targets three positions of the HCV genome.

Each shRNA pro-drug is expressed from the DNA template and processed to siRNAs by Dicer. The siRNAs associate with RISC, resulting in cleavage of HCV RNA. (Adapted from Lavender *et al.*, (2012))

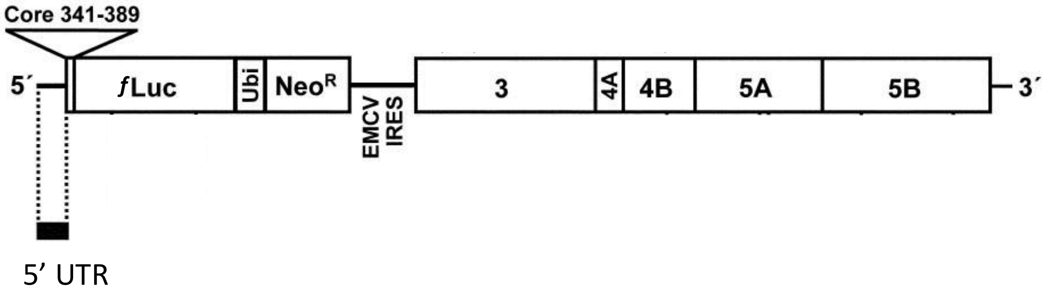


Figure 1.5 Schematic illustration of HCV I$_{389}$/NS3-3'/LucUbineo-ET subgenomic replicon.

The replicon is composed of the HCV 5' UTR , 342-389 nucleotides of the core region fused to a firefly luciferase gene (*f*Luc), the ubiquitin gene (Ubi), the neomycin phosphotransferase gene (Neo^{R}) which confers resistance to geneticin, the internal ribosome entry site (IRES) of encephalomycarditis virus (EMCV), then sequences for the non-structural HCV proteins NS3, NS4A, NS4B, NS5A and NS5B, and concludes with the HCV 3' UTR. The replicon is capable of persistent replication in Huh-7 cells. (Adapted from Krönke *et al.*, (2004))

1.2.4 HCV replicon reporter cell lines

The development of HCV replicon systems as a virology tool has greatly advanced HCV research (Berke et al., 2011). HCV replicon systems use a genetically modified HCV genome that incorporates a reporter signal and selectable marker (Figure 1.5). Replicon systems for many HCV subtypes have already been developed and utilized (Blight et al., 2003; Butt et al., 2011; Yu et al., 2014; Shier et al., 2016). HCV replicon systems are non-infectious and only express the non-structural proteins under control of a selectable marker. Although the replicon system does not produce all the virus life cycle phases (it does not have the entry and exit phases) it offers a system for drug activity validation. The non-structural proteins NS3 and NS5B are particular targets for drug development as they encode well-defined enzymatic activities crucial for virus replication (Blight et al., 2002). In some instances, adaptive mutations have been identified that increase viral replication in cell culture (Krieger et al., 2001; Lohmann et al., 2001).

1.3 Assays confirming RNAi activity

RNAi induced effects can be measured using a number of different indirect measurements. The reduction of target mRNA or protein compared to an untreated/non-specific control is the most commonly used approach in determining RNAi potency in cell culture systems, preclinical studies, as well as clinical trials. However, reduction of target levels is an implied proof of the expected RNAi mechanism of action, and other off-target processes may also drive the observed downregulation (Liang et al., 2013). Reverse-transcription quantitative PCR (RT-qPCR) is currently the method of choice for direct quantification of mRNA, as it is a sensitive, robust and well-established method. However, much care must be taken to avoid potential non-specific amplification and other artefacts. Holmes et al. (2010) and others (Shepard et al., 2005; Chen et al., 2011; Herbert et al., 2011) have shown that location of the primer pairs could have implications for validation of siRNA knockdown. For example, two different siRNAs were targeted to two locations on *PKC-ε* mRNA and then five primer pairs were used to amplify regions along the length of the mRNA. Knockdown of mRNA was detected with the primers targeting mRNA regions upstream of the target location, while primer sets designed to amplify targets downstream of the cleavage position, and particularly towards the poly A region of the mRNA were still able to detect mRNA fragments (Holmes et al., 2010). The 3' RISC cleaved product of targeted RNA is known to persist as it contains mechanisms (recognition elements) that bypass immediate degradation, allowing the 3' fragment to persistent after RISC cleavage (Orban and Izaurralde, 2005; Yoshikawa et al., 2013).

An indirect method of determining mRNA reduction is to measure the reduction of endogenous protein levels. However, this requires good, highly specific antibodies against endogenous proteins. Such antibodies are not always available and additionally, the method is not well suited to high throughput applications. Recombinant antigen tagged proteins (eg. HA, FLAG, His) can also be used, but remains an option only for validating assays in overexpression systems. Additionally, fluorescent or enzymatic reporters can be fused to gene targets to produce recombinant proteins for simplified, light-based target level quantification. Reporter constructs may also be generated by cloning the cDNA of the target downstream of the translational stop codon of the reporter. Transcription of the construct produces a chimeric mRNA from which only the reporter protein is translated. Targeting of the cDNA by RNAi leads to degradation of the entire mRNA and a proportional reduction in reporter is measured (Yeung et al., 2008; Lin et al., 2007). However, measurement of mRNA or protein by any of these methods is clearly insufficient evidence to conclude that the observed downregulation of a target is directly due to RNAi activity. Reduction of mRNA, and therefore protein may arise due to off-target effects, such as activation of Toll-like receptor (Robbins et al., 2009) which is induced by presence of siRNA (Grimson et al., 2007; Jackson and Linsley, 2010).

1.3.1 Assays for direct confirmation of RNAi activity

Detection and confirmation of the 5' end generated by AGO2-directed cleavage remains the only means of directly confirming RNAi activity. Initially, confirmation of RISC cleaved RNA used *in vivo* assays that relied on autoradiograpic detection of cleaved ^{32}P-cap-labelled synthetic RNA targets (Rand et al., 2005b; Koller et al., 2006). Adaptation of the 5' Rapid Amplification of cDNA Ends (5'RACE) assay originally developed to amplify full length mRNA transcripts (Frohman et al., 1988; Frohman, 1994; Schaefer, 1995) is currently the "gold standard" for confirming an RNAi mechanism of action (MOA) *in vivo* (Soutschek et al., 2004; Koller et al., 2006; Davis et al., 2010). In 5' RACE, an RNA adapter containing a 3' hydroxyl group is ligated to the 5' newly cleaved end of the target mRNA using T4 RNA ligase. The RNA adapter becomes a known sequence for the forward primer during PCR. A reverse primer specific to the transcript of interest is used in first strand cDNA synthesis. The cDNA product is then purified and PCR amplified using a forward primer specific to the 5' adapter and a reverse primer to the cDNA, resulting in a PCR product of defined size (Lasham et al., 2010). The method is commonly referred to as RNA ligase-mediated (RLM)-RACE (Figure 1.6). The amplicon is then sequenced (by Sanger sequencing) to confirm the position of the cleavage site.

Additionally, a number of authors have proposed modified protocols for confirming RISC cleaved products. For example, Lasham et al., (2010) proposed using a molecular beacon probe to specifically detect the junction between the RNA adapter and RISC-cleaved product of 5' RACE cDNA template or first round PCR reaction (Figure 1.6). However, although qPCR assays are now standard in many laboratories, each new target would require an optimised qPCR assay. Additionally, high throughput analysis would be hampered by upstream sample preparation activities such as RNA extraction, adapter ligation and RNA clean up by chloroform extraction. For the standard 5' RACE assay, fragment sequences are confirmed by cloning the PCR fragments into a vector and then sequencing a number of clones to identify the expected cleavage position. This is a time-consuming operation restricting its application in processing of large numbers of samples at one time.

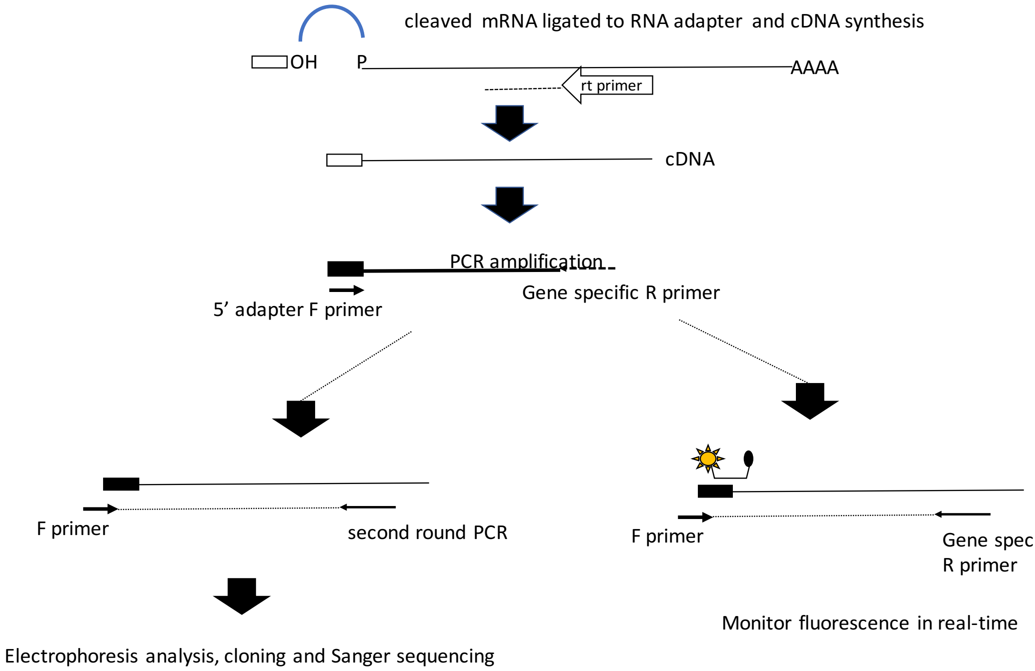

cleaved mRNA ligated to RNA adapter and cDNA synthesis
OH
P
AAAA
rt primer
cDNA
PCR amplification
5' adapter F primer
Gene specific R primer
F primer
second round PCR
F primer
Gene specific R primer
Monitor fluorescence in real-time
Electrophoresis analysis, cloning and Sanger sequencing

Figure 1.6 Diagram of RNA ligase-mediated Rapid Amplification of cDNA Ends (RLM RACE) vs molecular probe detection of 5' RACE products.

After targeted knockdown, total RNA is collected, adapter ligated and reverse transcribed using a gene specific primer. The cDNA is amplified and can then be cloned and sequences. Alternatively, the ligation position can be quantified by probe-based quantitative PCR.

1.3.2 Confirmation of RISC cleavage using RACE-Seq

In plants, miRNAs show extensive complementarity to their target RNA with many having perfect pairing (Baumberger and Baulcombe, 2005). These miRNAs are more appropriately termed endogenous siRNA and result in cleavage of their target. Degradome sequencing (also known as parallel analysis of RNA ends (PARE)) is a widely used method that reveals the activity of these endogenous siRNAs by identifying their targets. Degradome sequencing is a high throughput method for capturing uncapped RNA ends. An RNA adapter is directly ligated to the 5' end of uncapped RNAs prior to poly-A reverse transcription. The fragments are then digested by Mmel, a Type II restriction endonuclease. A 3' DNA adapter is added and the short fragments amplified and Next Generation Sequencing libraries prepared and sequenced (Addo-Quaye et al., 2008; German et al., 2008). The reads are then aligned to the transcriptome and compared to the plant miRNA database to reveal miRNA activity. Degradome sequencing has also been used to identify miRNA-dependant cleavage activity in humans (Bracken et al., 2011).

To date, there are only four publications applying (NGS) (*in lieu* of Sanger sequencing) of targeted 5' RACE products to confirm a specific interaction of introduced siRNA or shRNA via an RNAi MOA in humans or human cell lines (Tabernero et al., 2013; Denise et al., 2014; Barve et al., 2015; Ganesh et al., 2016). The assay is termed RACE-Seq and is very much in early implementation stage, with no standardized approach to performing and validating assays with variations in preparing NGS libraries, performing alignments, 5' end counting and data reporting. NGS is a powerful technology, with the capacity to bring new insights to RISC behaviour. Of the four publications, only Denise et al., (2014) sought to utilise RACE-Seq to infer RISC behaviour beyond confirming an expected cleavage product, with their result hypothesising that RISC may show preference for additional cleavage positions at nucleotides13-15.

A number of barriers currently hinder the fluid uptake of RACE-Seq application; these may include; the overall higher cost of NGS, long protocols involving manual handling operations, lack of experience in preparing RACE-Seq libraries, lack of experience of commercial NGS facilities in handling 5'RACE samples and lack of easy to use data analysis options.

1.4 Next Generation Sequencing: Ion Torrent PGM

NGS is also known as high-throughput sequencing, massive parallel sequencing or second-generation sequencing, and are terms used to describe different modern sequencing technologies such as Illumina sequencing, SOLiD sequencing and Ion Torrent sequencing. These sequencing technologies allow simultaneous sequencing of thousands to millions of individual DNA fragments. Steady improvements in technology, chemistry and data analysis improvement has seen the cost of NGS technologies decrease and the development of innovative sample preparation has led to expanded applications of NGS. The Ion Torrent sequencing platform differs substantially from the Illumina platform, although both systems require an amplified template signal, with the Ion Torrent system using clonal amplification on beads in an emulsion PCR reaction. The beads are then deposited into wells on the sequencing chip and the change in pH due to the release of a H^+ ion when a nucleotide is incorporated to a growing complementary strand is monitored for each well. This signal is interpreted and the number of bases incorporated determined.

1.4.1 NGS sample preparation and emulsion PCR for Ion Torrent Sequencing

NGS library preparation kits for Ion Torrent are now available from a number of manufacturers including; Thermo Fisher Scientific (who supply the official Ion Torrent validated protocols and reagents), New England Biolabs (NEB, Ipswich, MA), KAPA Biosystems (Roche), BIOO Scientific (Austin, TX) and LABGENE Scientific SA (Freiburg, Switzerland). Sample and reagent storage as well as good laboratory practices are particularly important, with the use of precision pipettes, filter pipette tips, and certified nuclease-free consumables highly recommended. There are a number of standardized sample requirements when preparing NGS libraries. It is recommended to assess DNA samples by gel electrophoresis to confirm samples are of the expected size and that DNA is of satisfactory integrity. The quality of the purified DNA should yield an$OD_{260/280}$ ratio of 1.8 to 2.0 and an $OD_{260/230}$ ratio of 2.0 to 2.2, the former being more critical. Samples may be rejected for sequencing based on poor OD assessment, or, poor OD assessment may indicate that the sample requires a clean-up procedure before input to library preparation (Endrullat et al., 2016). The sample amount needs to be accurately assessed, a step that becomes particularly important when processing multiple samples. Accurate DNA quantification is carried out using fluorescent-based DNA quantification methods such as the Qubit DNA assay (Thermo Fisher Scientific) or using electronic fragment analysers such as the Agilent Bioanalyzer system (Agilent Technologies, Santa Clara, CA).

For Ion Torrent library preparation, double stranded DNA is blunt ended and 5' phosphorylated using a mixture of enzymes (Head et al., 2014). A clean-up step may be required at this point for some protocols. The Ion Torrent Sequencing adapters (termed A-adapter and P1-adapter) are then added to the blunt ended fragments in a ligation reaction. The A-adapter has a short barcode allowing samples to be individually tagged for later pooling and sequencing in a single reaction (multiplex sequencing), and also holds the consensus key, which is a short sequence used to validate the sequencing signal at the start of the sequencing reaction. The P1 adapter sequence is the end that becomes attached to the Ion sphere particles during emulsion PCR. Since the fragment ends and adapter ends are both blunt ended, the adapters ligate to the fragment ends in a mix of configurations (Figure 1.7). The adapter ligated samples undergo a bead-based clean-up protocol. DNA clean-up using magnetic beads can efficiently remove free adapter and short fragments from sample preparations (Bronner et al., 2013). A short round of PCR is commonly used to enrich for fragments that have P1-adapter on one end and A-adapter on the other. Another round of bead clean-up is then performed. An alternative approach is to quantify the library using quantitative PCR. Fragments having a P1and A-adapter are amplified and a region on the P1 adapter is used for Taqman probe hybridisation, thus only fragments that have the potential to be amplified in emulsion PCR are detected and quantified.

The Ion Torrent platform requires that the final DNA library is within the validated fragment size range, thus additional size selection may be required prior to sequencing. A number of manual and automated protocols and instruments are available for size selection of DNA fragments. These may include: manual extraction by gel excision and purification, E-gel Size Select system (Thermo Fisher Scientific) and automatic size selection instruments such as the LabChip XT (PerkinElmer Inc., MA, USA) and Pippin Prep (Sage Science Inc., MA, USA). Irrespective of method used, validation of fragment size and accurate quantification are critical for input to emulsion PCR.

The OneTouch^{TM}2.0 emulsion PCR system is a fully integrated automated system that performs the clonal amplification of the DNA library onto the sequencing beads. The instrument generates an emulsion consisting of an aqueous phase in oil. The aqueous phase has all the components for PCR, ie. template (ideally a single fragment of the

DNA library per droplet), amplification primer, bead (the bead holds a primer tag complementary to the P1-adapter sequence), DNA polymerase, dNTPs and required buffer reagents for PCR (Figure 1.8). This bead emulsion undergoes *in situ* thermal cycling PCR and finishes with breaking of the emulsion using detergent, allowing the beads to be recovered by centrifugation (Merriman et al., 2012). The extracted beads are now templated with the library, but need to undergo further enrichment to select for beads that have been well templated and to exclude 'empty' beads. This is achieved using the OneTouch Enrichment System which is also a fully integrated system. First, a biotinylated tag is hybridized to the terminal adapter sequence of the templated bead pool. Beads that did not template will not be captured. The biotin-tagged beads are then bound to magnetic streptavidin capture beads and the complex is collected magnetically, washed several times and the templated beads eluted. This eluate is the final templated beads that will be prepared for sequencing (Kohn et al., 2013). Validated chemistries, protocols and instrument calibration ensure that the emulsion PCR and enrichment processes typically yield over 98% of a bead population carrying template.

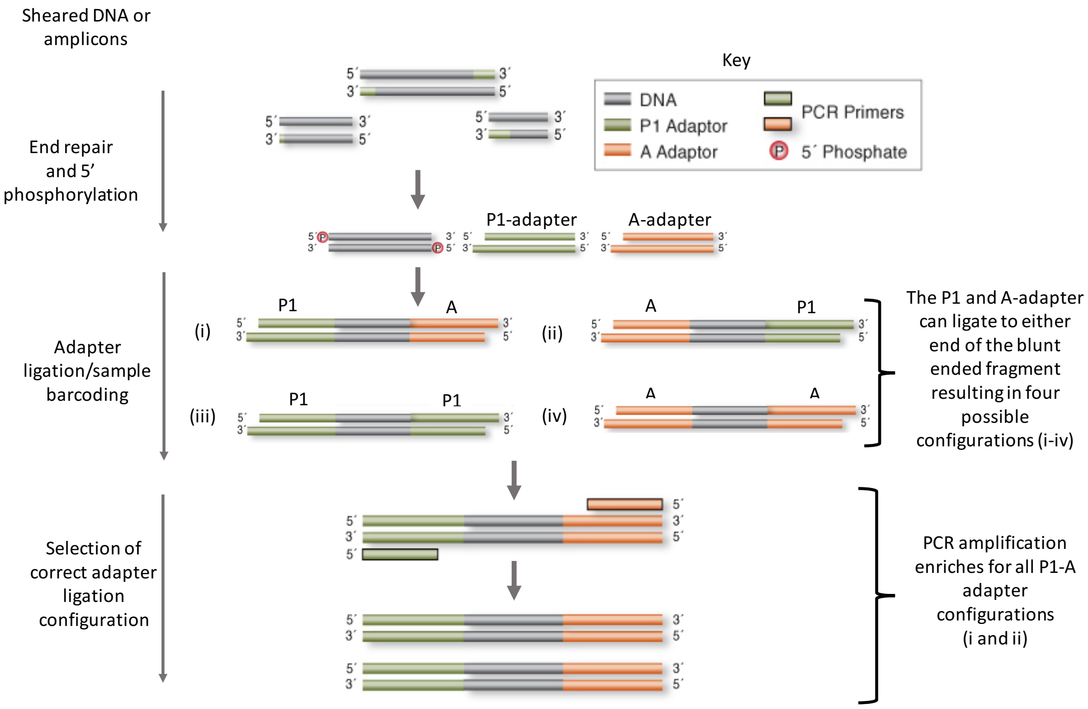

Sheared DNA or amplicons
End repair and 5’ phosphorylation
Adapter ligation/sample barcoding
Selection of correct adapter ligation configuration
Key
DNA
P1 Adaptor
A Adaptor
PCR Primers
5´ Phosphate
P1-adapter
A-adapter
(i)
(ii)
(iii)
(iv)
P1
A
5´
3´
The P1 and A-adapter can ligate to either end of the blunt ended fragment resulting in four possible configurations (i-iv)
PCR amplification enriches for all P1-A adapter configurations (i and ii)

Figure 1.7 Workflow of library preparation for the Ion Torrent Sequencing Technology.

Fragmented DNA or PCR products undergo end repair, which generates blunt ends. The ends are 5’ phosphorylated and can be tagged with the sequencing adapters. The fragment ends are ligated with either the A adapter or the P1 adapter. A short round of PCR selects for the correct orientation of adapters to generate the final NGS library. (Adapted from: https://www.neb.com/products/e6270-nebnext-fast-dna-library-prep-set-for-iontorrent)

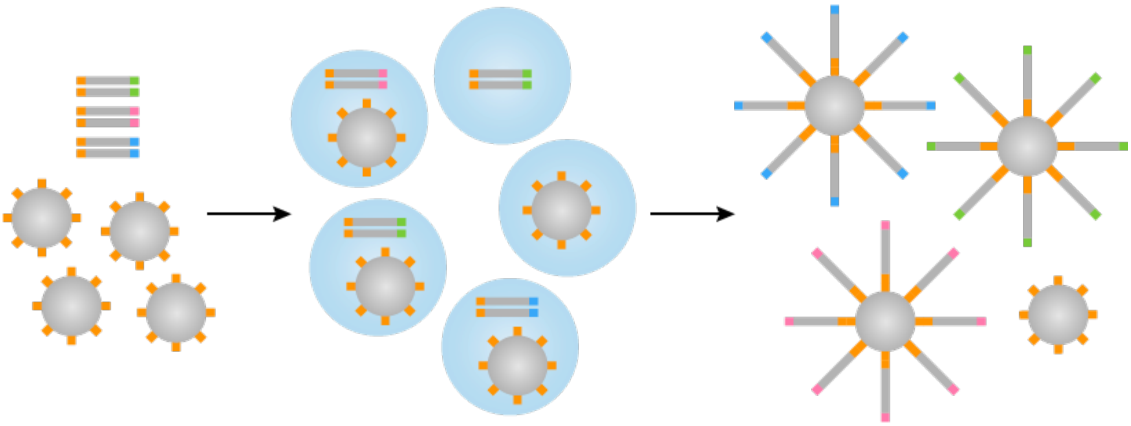


Figure 1.8 Illustration of Emulsion PCR reaction utilised in Ion Torrent library preparation.

The OneTouch instrument preforms the emulsion PCR reaction. A mix of emulsion oil, sequencing beads tagged with a primer complementary to the P1 adapter sequence, PCR mix, primer and library DNA are emulsified to create micro-reactors. PCR cycling denatures the double stranded DNA and one strand is templated to the bead. PCR cycling generates copies of the DNA in each of the reactors. The free strand is also copied, increasing the number of templates attaching to the beads. (Adapted from: http://www.atdbio.com/content/58/Next-generation-sequencing#figure-emulsion-pcr)

1.4.2 Ion Torrent semiconductor sequencing technology

The Ion Torrent Technology uses a semiconductor sensor array chip for sequencing and data collection with a simple, on-chip sequencing chemistry (Figure 1.9). Templated beads from emulsion PCR are loaded onto the sequencing chip, where one templated bead occupies a single well. The chip is a highly sensitive pH meter. Using native dNTPs, one nucleotide at a time is flowed into the wells. Incorporation of dNTPs in a template specific manner releases hydrogen ions (H^+) from the 3' OH^- incorporation site on the growing strand. The change in pH is measured and converted to an electric signal. Unincorporated nucleotides are washed away and the next nucleotide added. Importantly, the chemistry is such that more than one nucleotide can be added to the template in a single flow, e.g. due to homodimer, homotrimer etc. occurrence (i.e. TTT, AAA). The signal is then averaged and the number of incorporated nucleotides determined (Rothberg et al., 2011). The base-calling software then converts the raw data to a nucleotide sequence.

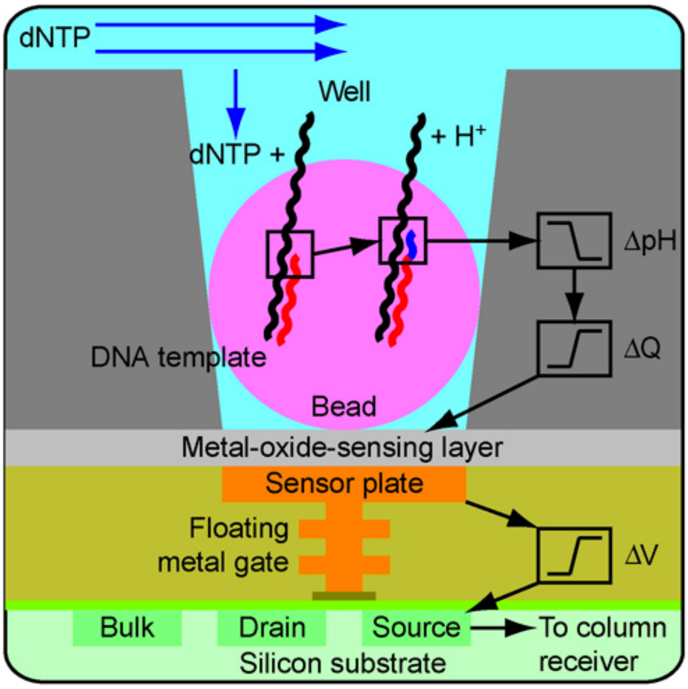


Figure 1.9 Illustration of the Ion Torrent Semiconductor chip based sequencing technology.

Detection of incorporated nucleotides is correlated with a change in pH from the release of H+ at each flow of nucleotide. (Adapted from: Rothberg et al., 2011)

1.4.3 On-system data processing and analysis

When the base-call analysis has completed, data quality is assessed by the Ion PGM Torrent Server. Proprietary algorithms filter the base-called data for quality assurance and removes any 'obviously low quality' reads. Low quality removal includes; the removal of reads that have fallen out of phase (i.e. some fragments being sequenced may fall behind cycle phase due to improper extension or may advance ahead of phase due to carryover of reagents from the previous nucleotide flow), reads with 'signal droop' (which is when the signal strength of some reads becomes weaker over the course of the sequencing reaction, and this 'signal droop' can be detected, quantified and used as a criteria for eliminating reads from the final dataset), detection and elimination of reads with well-to-well crosstalk, and low clonal intensity (the signal from each well can be assessed and the relative number of templates that were primed on each bead determined) (Merriman et al., 2012). After this pre-filtering for quality, reads are de-multiplexed, any reads that are adapter-adapter ligations (adapter dimer) or polyclonal (where a mix of templates became templated to the same bead) are removed. A report summarising chip loading efficiency, filtered read statistics, de-multiplexed sample information (barcode ID, sample name, number of reads per sample and average length of fragments per sample) is obtained. Data can be made available in a number of formats such as FASTQ and BAM file formats. A number of additional data analysis protocols can be performed directly on the system such as FASTQC Quality assessment, alignment to a reference genome, variant analysis and *de novo* genome assembly (Seo et al., 2015).

1.5 A brief overview of Illumina Sequencing Technology

The two major players in benchtop short read sequencing platforms are Illumina and Ion Torrent. The Illumina sequencing platform remains the most widely used sequencing technology. Illumina sequencing uses clonal amplification on a solid surface and reversible terminator technology (Figure 1.10). After sample preparation, library fragments are attached to the sequencing flow cell. The flow cell is covered with two different types of oligo tags that are complementary to the ends of the library fragments. The fragments hybridize to the tags and become templates for synthesis of a complementary fragment that is now attached to the flow cell. The original fragment is washed away and clonal amplification of each fragment occurs by bridge amplification by the attached fragment folding over and hybridizing to a neighbouring tag and a complementary fragment is synthesised resulting in a copy of the DNA strand. This reaction occurs a number of times to create up to 1000 identical clones in close proximity. The reverse strands are then removed, leaving a single strand for templating. A sequencing primer is added and the forward strand sequenced (Head et al., 2014).

Illumina technology uses fluorescently labelled dNTPs with bound reversible terminators. During sequencing, all 4 nucleotides are accessible to the bound clusters, but due to the terminator, only one nucleotide can attach to the growing complementary strand. Unbound nucleotides are washed away and incorporated nucleotides are stimulated by a laser with each fluorescently-tagged nucleotide emitting a characteristic signal. Each strand in the cluster emits the same signal, providing an amplified signal that is optically captured. The terminator is enzymatically cleaved allowing the next base of the template to be incorporated (Figure 1.10). This process of Illumina sequencing is commonly referred to as sequencing by synthesis. Since one nucleotide is incorporated at a time, the number of flows dictates the read length. Illumina Technology employs paired-end sequencing such that each read can be sequenced in both directions. Thus, for short reads (up to 250 bp) for current chemistries the reads obtained in both directions helps to eliminate sequencing errors, as the read pairs should match. This approach is highly effective in resolving structural re-arrangements such as insertions, deletions and inversions. In the case of long templates, a portion of the strand is read in each direction, a strategy which is advantageous for the assembly of whole genome reads or for resolving repetitive genomic regions compared to single end sequencing (Treangen and Salzberg, 2011). For paired-end sequencing, on completion of the first-strand sequencing reaction, the cluster ends anneal to the second tag on the flow cell, and the second index sequence is read and the strands extended using the first strand as template. This process generates clusters with paired strands. The forward strand is then removed with the remaining strand becoming template for sequencing (Ansorge, 2009).

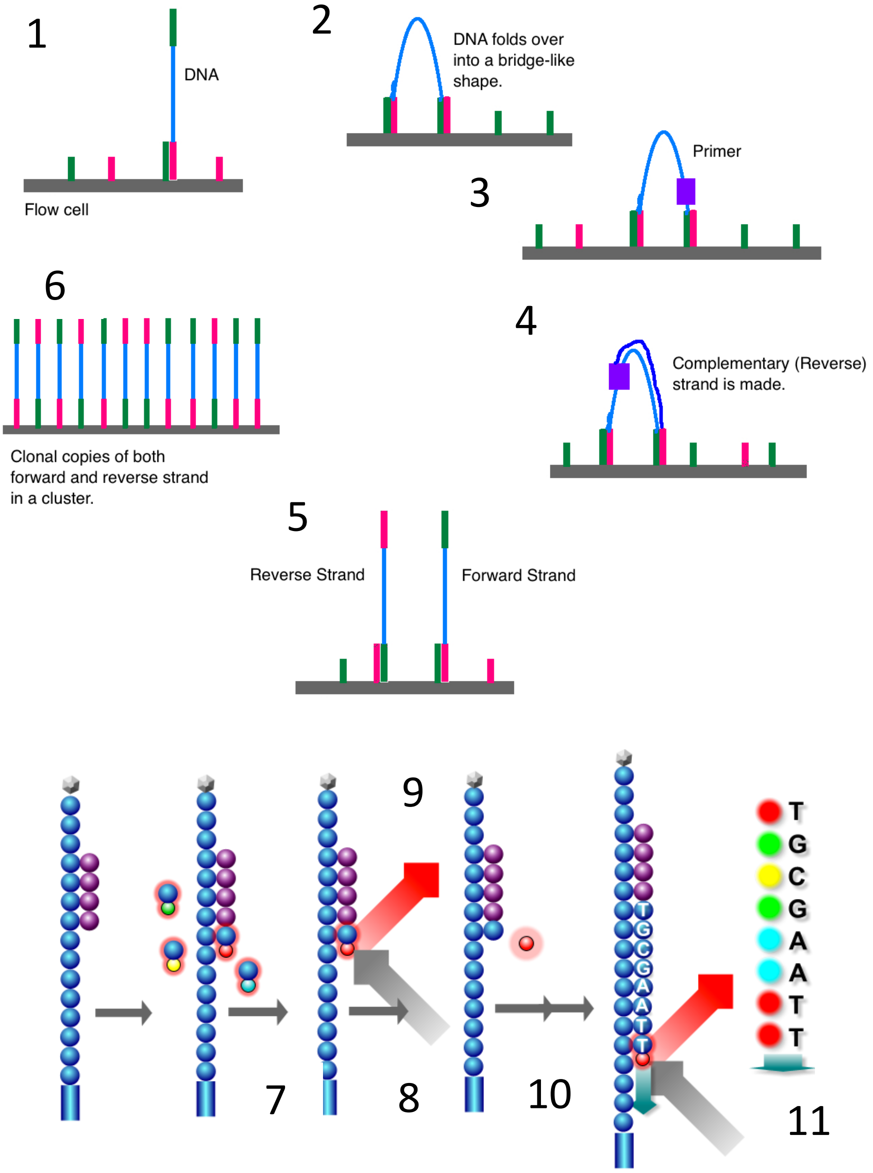

1
DNA
Flow cell
2
DNA folds over into a bridge-like shape.
Primer
3
6
4
Complementary (Reverse) strand is made.
Clonal copies of both forward and reverse strand in a cluster.
5
Reverse Strand
Forward Strand
9
7
8
10
T
G
C
G
A
A
T
T
11

Figure 1.10 Outline of Illumina sequencing process.

Adapter ligated fragments are spread out and captured to a primer-loaded flow cell (1), the template strand folds over (2), and is cloned by bridge PCR amplification (3 and 4) to produce two strands (5). This continues for a number of cycles to produce clusters of fragments (6). After addition of the sequencing primer, a labelled dNTP is added to the growing strand (7). The flow cell is excited by a laser (8) and the fluorophore emission captured by a camera (9). The sequencing terminator is removed (10) and the sequencing cycle continues with the addition of the next nucleotide. The base-calling software decodes the raw data generating the sequence (11). (Adapted from https://commons.wikimedia.org/wiki/File:Cluster_Generation.png, and https://en. wikipedia.orgwiki/File:Sequencing_by_synthesis_Reversible_terminators.png)

1.6 Aims of the project

The current dogma of RNAi stipulates that active AGO2 RISC cleaves the target RNA between bases 10-11 when counting from the 5' end of the guide strand of the introduced RNAi modality (Martinez and Tuschl, 2004; Foster et al., 2012). The 5' RACE-PCR assay remains the only means for directly confirming AGO2-RISC cleavage activity. The assay is poorly reported in pre-clinical and clinical trials. This is likely due to its low throughput capacity and the need for extensive manual preparations; as a result, single instances of an on-target MOA are extrapolated as continued evidence across drug development pipelines. Sanger sequencing is used to confirm the expected cleavage position. Transitioning the assay to Next Generation Sequencing is likely to improve throughput and provide opportunities for greater insight to RISC activity.

1.7 Work presented in this thesis

This work represents a critical evaluation of the application of NGS to 5'RACE analysis of RISC cleaved transcripts and opportunities for improving the technique. Briefly, the aims of the project were:

Technical aspects:

1. To transition the Illumina HiSeq RACE-Seq protocol (Denise et al., 2014) to the Ion Torrent PGM.

2. To improve the 5' RACE-Seq assay to allow multiplexing of up to 10 samples per Ion Torrent 318 sequencing chip.

3. To evaluate the NGS library preparation criteria required to generate sufficient RACE-Seq data using the Ion Torrent PGM.

Data analysis:

4. To evaluate and implement a simple data analysis workflow for non-expert analysis of RACE-Seq data.

5. To investigate whether RACE-Seq data provides insights into RISC cleavage beyond the expected cleavage behaviour.

Biological insights:

6. To investigate the RISC cleavage activity of siRNAs, shRNAs and a Dicer substrate siRNA (DsiRNA).

2 METHODS AND MATERIALS

2.1 RNAi-based bioactive oligonucleotides

The sequences for all six RNAi oligonucleotides targeting the HCV replicon genome were as reported by Lavender et al. (2012) and Denise et al. (2014). The three synthetic shRNA analogues (Table 2.1) were designed to mimic the shRNA prodrug sequences expressed from TT-034, a DNA-directed RNAi-based drug that expresses three shRNAs targeting three different locations on the HCV genome (Figure 1.4). The three siRNA synthetic analogues were designed as the most prominent sequences processed by Dicer for the three shRNAs expressed from TT-034 (Denise et al., 2014).

Table 2.1 Bioactive oligonucleotides targeting against HCV replicon genome

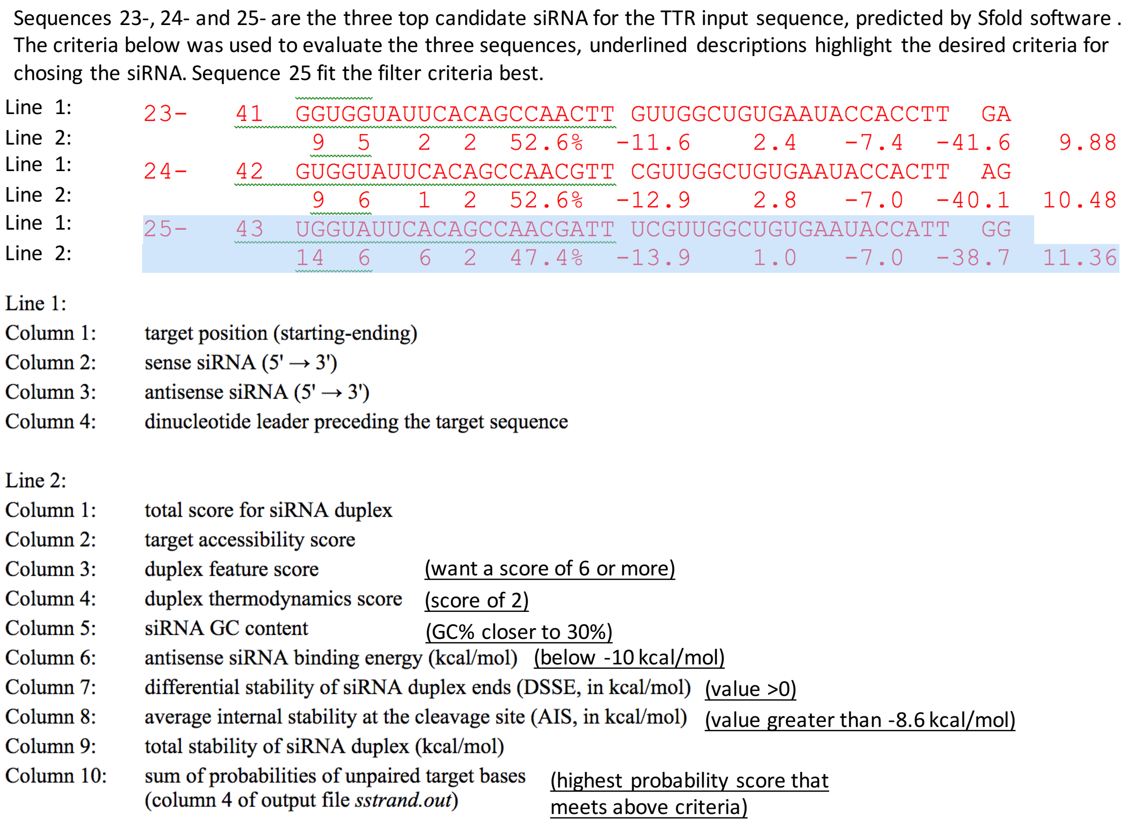
	Guide strand	Passenger strand
siRNA22	5'-AUUGGAGUGAGUUUAAGCUga-3'	5'-AGCUCAAACUCACUCCAAUuu-3'
siRNA19	5'-CAACUCCUGGCUAGGCAAuuugu-3'	5'-uagUUGCCUAGCCAGGAGUUGAC-3'
siRNA6	5'-GAAAGGCCUUGUGGUACUgaa-3	5'-gAGUACCACAAGGCCUUUCGC-3'
shRNA22	5'-AUUGGAGUGAGUUUAAGCUgaagcuu**g**AGCUUAAACUCACUCCAAUuuuuu-3'	
shRNA19	5'-GUCAACUCCUGGCUAGGCAAuuuguguagUUGCCUAGCCAGGAGUUGACuuuuuu-3'	
shRNA6	5'-CGCGAAAGGCCUUGUGGUACUgaagcuugAGUACCACAAGGCCUUUCGCuuuuu-3'	

For shRNA: lowercase = hairpin loop structure
For siRNA: The siRNAs were selected as the most prominent maturation products from shRNA6, shRNA19 and shRNA22 expression from TT-034 (Denise et al., 2014). As a consequence, all of the siRNAs hold part of the shRNA loop sequence.

2.2 Rational design of a siRNA targeting Transthyretin

The widespread use of siRNAs in genomic studies has resulted in a variety of online tools that incorporate the many design guidelines for generating highly effective small interfering RNA sequences. The online software package *S*fold v2.2 (Software for Statistical Folding of Nucleotides and Studies of Regulatory RNAs) web service (http://sfold.wadsworth.org/cgi-bin/sirna.pl) was chosen for its simple application and interpretation of results. The software incorporates methodologies for siRNA design, including target accessibility prediction and siRNA duplex thermodynamic properties and generates both graphical and text output (Ding et al., 2004). An easy to follow filtering and scoring system makes this a very user-friendly software package.

To design the siRNA sequence, a 100 bp region within the coding sequence of Transthyretin (TTR*)* mRNA (NM_000371.3) from position 450 bp to 550 bp was chosen as input into the *S*irna programme within *S*fold v2.2 web service and siRNA analysis executed using the default design criteria. In order to view the siRNA sequences generated, the "Output for all siRNAs" option was selected. This produced an analysed text output showing a list of all possible siRNAs, beginning at the start position of the input sequence. An example of the analysed sequence is presented in Figure 2.1, which also highlights the filtering and scoring criteria that was used to choose siRNA-TTR. A duplex thermodynamic score of 2, was identified as an ideal start point for filtering the initial siRNA sequence, as this score is derived from two other design criteria namely; 1 point is gained for siRNAs with differential stability of siRNA duplex ends with score >0 and 1 point is gained for siRNAs with average internal stability values >-8.6 kcal/mol. A requirement of siRNA sequences to meet a duplex thermodynamic score of 2 filtered the siRNA list from 106 possible sequences to 3 sequences. These three siRNA sequences were evaluated using the pass criteria in Figure 2.1. The sequence with the most pass criteria was selected. BlastN analysis confirmed that the siRNA sequence selected was specific for *TTR* mRNA. The chosen siRNA sequence was synthesised by IDT, as individual RNA strands, siRNA-TTR-AS 5'-UCGUUGGCUGUGAAUACCAtt-3'and siRNA-TTR-SS 5'-UGGUAUUCACAGCCAACGAtt-3' (at 100 nmol with standard desalting purification, capital letters=RNA bases, tt=DNA bases, AS = antisense strand, SS = sense strand).



Sequences 23-, 24- and 25- are the three top candidate siRNA for the TTR input sequence, predicted by Sfold software . The criteria below was used to evaluate the three sequences, underlined descriptions highlight the desired criteria for chosing the siRNA. Sequence 25 fit the filter criteria best.

```
Line 1:    23-    41  GGUGGUAUUCACAGCCAACTT GUUGGCUGUGAAUACCACCTT  GA
Line 2:                 9   5    2   2  52.6%  -11.6     2.4    -7.4   -41.6   9.88
Line 1:    24-    42  GUGGUAUUCACAGCCAACGTT CGUUGGCUGUGAAUACCACTT  AG
Line 2:                 9   6    1   2  52.6%  -12.9     2.8    -7.0   -40.1  10.48
Line 1:    25-    43  UGGUAUUCACAGCCAACGATT UCGUUGGCUGUGAAUACCATT  GG
Line 2:               14   6    6   2  47.4%  -13.9     1.0    -7.0   -38.7  11.36
```

Line 1:

Column 1: target position (starting-ending)

Column 2: sense siRNA (5' → 3')

Column 3: antisense siRNA (5' → 3')

Column 4: dinucleotide leader preceding the target sequence

Line 2:

Column 1: total score for siRNA duplex

Column 2: target accessibility score

Column 3: duplex feature score (want a score of 6 or more)

Column 4: duplex thermodynamics score (score of 2)

Column 5: siRNA GC content (GC% closer to 30%)

Column 6: antisense siRNA binding energy (kcal/mol) (below -10 kcal/mol)

Column 7: differential stability of siRNA duplex ends (DSSE, in kcal/mol) (value >0)

Column 8: average internal stability at the cleavage site (AIS, in kcal/mol) (value greater than -8.6 kcal/mol)

Column 9: total stability of siRNA duplex (kcal/mol)

Column 10: sum of probabilities of unpaired target bases (column 4 of output file *sstrand.out*) (highest probability score that meets above criteria)

Figure 2.1 Criteria for selection of siRNA-TTR.
A 100 bp region of the *TTR* mRNA sequence (NM_000371.3) from position 450 – 550 bp, was used to select an siRNA using *S*fold. Initially, the three siRNA sequences (in red) were identified with a duplex thermodynamic score of 2, and these were then further assessed against the selection criteria under lane 2. The minimum pass criteria are underlined. The blue highlighted sequence passed the selection criteria.

2.3 Design and synthesis of a Dicer Substrate siRNA targeting Transthyretin

A custom Dicer substrate siRNA (DsiRNA) was selected using the IDT RNAi Design Tool (https://eu.idtdna.com/site/order/designtool/index/ DSIRNA_PREDESIGN). The same 100 bp region of the *TTR* mRNA sequence (position 450 bp to 550bp) used to design the siRNA was input into the RNAi design tool to generate a list of DsiRNAs. The tool automatically aligns the DsiRNAs to the Human transcriptome. From the list generated, one molecule, ID: CD.Ri.25599.13.23, was specific for *TTR* mRNA only, while the other molecules generated from the list showed alignment to various human transcripts. DsiRNA-TTR was synthesised by IDT at 2 nmol with affinity purification and was supplied as a pre-annealed lyophilized sample. The sequence for DsiRNA-TTR was sense strand 5'-AUGCAGAGGUGGUAU UCACAGCCaa-3' and antisense strand 5'-UUGGCUGUGAAUACCACCUCUGCAUGC-3' (capital letters=RNA bases, aa = DNA bases).

2.4 Primer design

A truncated GeneRacer adapter was used for RLM RACE assays. The RNA adapter sequence was as reported in Denise et al., (2014) (Table 2.2). All primer sequences used in RLM RACE assays for HCV and TTR are listed in Table 2.3. Primers were designed using the Integrated DNA Technologies (IDT) PrimerQuest Tool (https://www.idtdna.com/PrimerQuest/Home/Index?Display=AdvancedParams) and analysed using the Oligoanalyzer 3.1 Tool (https://eu.idtdna.com/calc/analyzer), both from IDT. The 5' RACE primers were designed to ideally to be length 22-28 nucleotides, have GC content of 50-70%, and no more than two G or C residues at the 3' end of the primer. Primers were analysed for self-complementarity and hairpin formation. The gene-specific reverse primers were designed with the same criteria. PCR amplification primers were additionally designed to have similar melt temperature to the GeneRacerF1 primer. The annealing temperature for amplification primer pairs was checked using the NEB Tm Calculator (https://tmcalculator.neb.com/#!/). Primers were synthesised by IDT at 25 nmol with standard desalting. The quantitative PCR (qPCR) amplification primers for measuring knockdown of *TTR* mRNA were as reported in Hayashi et al. (2012) (Table 2.3). GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was used as the internal reference control (see Table 2.3 for primer sequences).

Table 2.2 GeneRacer RNA adapter sequence

Purpose	Name	Sequence
Adapter ligation	GeneRacer adapter	GGACACUGACAUGGACUGAAGGAGUAGAAA

Table 2.3 RACE-Seq primers used in this thesis

Purpose	Name	Sequence	Melt Temp (^{o}C)	Expected size of RACE-Seq amplicon
reverse transcription/PCR	22+2*	CGAACCAGCT GGATAAATCCAACTGGGA	62	78 bp
reverse transcription	6+L1	GTATCTCTTCATAGCCTTATGCAGTTG	55.3	
PCR	6+L3	GCCGGGCCTTTCTTTATGT	55.7	182 bp
PCR	6+L2	GCGCCC GTT GGT GTT ACG	59.6	144 bp
reverse transcription	19+L4	GGCCATGGAGTCGTTGAA	55.2	
PCR	19+L2	GTAGGTCAAGTGGCTCAATGGAGTA	58.1	186 bp
PCR	19+L1	TTGTTCCTGAGCTAGAAGGATGGAGAAG	59.4	121 bp
reverse transcription primer	22+L2	TAGGAGTAGGCACCACATGAA	55.2	
PCR	22+L1	GGTCGGGCACGAGACAG	58.6	126 bp
PCR	22+L2	GACAGGCTGTGATATATGTC	50	114 bp
reverse transcription primer	TTR+L1	TTGTCTCTGCCTGGACTTCTAACATAGC	59.7	
PCR	TTR+L2	TCGTCCTTCAGGTCCACTGGAGGAGAAGT	65	~170 bp
qPCR	human TTR forward	CATTCTTGGCAGGATGGCTTC	56.5	
qPCR	human TTR reverse	CTCCCAGGTGTCATCAGCAG	57.7	
qPCR	GAPDH forward	CCCACTCCTCCACCTTTGACG	59.9	
qPCR	GAPDH reverse	CCACCACCCTG TTGCTGTAG	58.1	
PCR	GeneRacerF1	GGACACTGACATGGACTGAAGGAGTA	59.6	
PCR	GeneRacerF2	GGACACTGACATGGACTGAAGG	57.4	

*Primer selected from Denise et al., 2014

2.5 Cell culture and Bioassays

2.5.1 Routine cell maintenance

Experiments were carried out on a human hepatocellular carcinoma cell (Huh-7) line which harbours the I389/NS3-3'/LucUbiNeo-ER replicon of HCV genotype 1b (Con1 isolate) (Frese et al., 2002) kindly gifted from ReBLikon GmbH, Schriesheim, Germany). The cell line is commonly referred to as Huh7/Con1b. The cells expressing this non-packaging, self-replicating, luciferase-expressing subgenomic replicon of HCV were maintained as an adherent monolayer in 75 cm^{2} flasks and propagated in complete media. At 80-90% confluency spent media was poured removed and cells washed for 1 minute with 5 ml of pre-warmed Dulbecco's Phosphate-Buffered Saline(DPBS) (Thermo Fisher Scientific, Waltham, Ma) to remove residual media serum. Cells were detached in 1-2 ml of TrypLE Express reagent (Thermo Fisher Scientific) and incubated at 37^{o}C for 4-5 minutes. Flasks were observed under the microscope to ensure sufficient cell detachment and dislodged by gentle but firm tapping of the flask. TrypLE Express reagent was deactivated by washing cells with 5 ml of pre-warmed complete media. Cells were centrifuged in 15 ml conical tubes at 400 rpm for 5 minutes at 18°C, and the resulting pellet re-suspended in 3-9 ml of fresh media. Cells were split 1:3 or 1:4 as required and incubated at 37°C with 5% CO_2. Cell passage was normally carried out 2-3 times per week. For HCV knockdown experiments, cells at up to passage 38 were used.

The complete media consisted of:

Dulbecco's Modified Eagle's Medium, (DMEM) (Thermo Fisher Scientific) supplemented with 10% (w/v) foetal bovine serum (FBS) (Thermo Fisher Scientific), 1 mM sodium pyruvate (Thermo Fisher Scientific), 1x non-essential amino acids (Thermo Fisher Scientific, UK), 1x penicillin-streptomycin (Thermo Fisher Scientific) and 0.5 mg/ml Geneticin (Thermo Fisher Scientific).

2.5.2 Cell revival and plating

Stored cells were removed from liquid nitrogen and thawed in a 37°C water bath for 2 minutes with 2-3 inversions during thawing. Cells were added to 10 ml of pre-warmed complete media without Geneticin and were gently pipetted up and down to mix. The entire volume was transferred to a 25 cm^2 flask and incubated overnight at 37°C in a humidified 5% CO_2 incubator. The following day, media was changed to DMEM complete media (as above) containing 0.5 mg/ml Geneticin. At 80-90% confluence, cells were transferred to a 75 cm^2 flask for routine cell maintenance.

2.5.3 Cell counting

For experimental purposes, cells were counted and viability determined using Trypan Blue dye (Thermo Fisher Scientific) (Riss et al., 2004). Cells were detached and pelleted as previously described and re-suspended in 3-5 ml complete media or Opti-MEM reduced serum media (Thermo Fisher Scientific). Cells were mixed thoroughly by gentle pipetting at least 5 times using a 10 ml serological pipette and immediately, 100 µl of cell suspension was diluted with 100 µl Trypan Blue for a 1:2 dilution or a mix of 100µl Trypan Blue and 200 µl DPBS for a 1:4 dilution. Then, 10 µl of cell suspension in Trypan Blue was added to each side of a haemocytometer chamber and cells counted under an inverted phase-contrast microscope.

The number of total cells vs live cells/ml was established using the formula:
Average count x n x 10^4 = number of cells/ml

Where n = dilution factor and the average count is the total number of cells divided by the number of counting chamber squares, and where 10^4 is the multiplication factor to convert chamber depth (0.1 mm^3) to cm^3

2.5.4 Cell storage

Single or multiple 75 cm^2 flasks of confluent (80-90%) cells were washed with DPBS and detached with TrypLE Express reagent and cells pelleted as described in 2.3.1. Cells were re-suspended in 5 ml of freezing medium consisting of 5 ml DMEM, 20% FBS and 10% DMSO). Cells were aliquoted at 1 ml volumes to cryogenic vials and kept overnight at -80°C before being transferred to liquid nitrogen storage.

2.6 RNAi knockdown assays

2.6.1 Dose response assays

Oligonucleotides were tested at half-logarithmic dilutions for dose response assays. The concentration of active RNAi oligonucleotide ranged from either 5 nM to 0.167 fM with a final oligonucleotide concentration of 30 nM for siRNA22 and shRNA22 and 50 nM to 0.16 pM for siRNA6, shRNA6, siRNA19 and shRNA19, with a final oligonucleotide concentration of 50 nM. The final concentration of bioactive RNAi analogue was therefore as follows: 50 nM, 15.9 nM, 5 nM, 1.6 nM, 0.5 nM, 0.16 nM, 0.05 nM, 0.016 nM, 0.005 nm, 0.0016 nM, 0.0005 nM and 0.00016 nM.

All stock oligonucleotides were first diluted from 100 µM concentration to 5 µM working stock concentrations in nuclease-free water. For dose-response assays working dilutions of the RNAi oligonucleotides were first prepared at 2000 nM, 100 nM, 3 nM and 0.1 nM dilutions in Opti-MEM reduced serum media. To obtain a final concentration range, each of the dilutions were prepared at 2X concentration so that after addition of the cells to the assay plate, the final concentration of active RNAi reagent is at the desired concentration. Working in a deep-well plate, the assay dilutions were prepared by adding the required volume of oligonucleotide (to give 2x the desired concentration) to assay wells. The final concentration of oligonucleotide for each dilution was equalised to 30 nM or 100 nM using the non-specific oligonucleotide (NSO) (5’-GACCACTTGCCACCCATC-3’) and the volume made up to 20 µl volume. For the NSO control, NSO was diluted to 30 nM or 100 nM in 20 µl Opti-MEM. The mock control consisted of diluted transfection reagent in Opti-MEM. The non-transfected control consisted of Opti-MEM only.

Dharmafect3 (Thermo Fisher Scientific) was diluted in Opti-MEM to 0.0025 µl/µl by adding 24 µl of Dharmafect3 to 9.6 ml of Opti-MEM. The Dharmafect3 dilution was mixed thoroughly and 330 µl added to each deepwell oligo dilution, to give a total reaction volume of 350 µl. This reaction was mixed and 50 µl of each dilution transferred to assay plates in triplicate. Assay plates consisted of one white 96-well plate (VWR, Leicestershire, UK) for the luciferase assay and one clear cell culture plate (Corning) for the MTT assay. The NSO control preparation had 480 µl diluted Dharmafect3 added to it and 50 µl volumes transferred to the assay plate in triplicate. For the mock control, 50 µl of diluted Dharmafect3 was added to assay mock control wells. The non-transfection control wells had 50 µl Opti-MEM added to them.

The Dharmafect3-oligo complexes were allowed to form for 30-45 minutes during which time the cells were prepared. Cells from two or three 75 cm^2 flasks at 80-90% confluency were detached and pelleted as per section 2.5.1. Cells were re-suspended in 5 ml of Opti-MEM, counted in duplicate by Trypan Blue assay, and diluted as necessary with Opti-MEM to obtain 7000 cells per 50 µl volume. At the end of the oligo-complex formation incubation step, 50 µl of cell suspension was added to each well. Plates were viewed under the microscope to ensure even distribution of cell in the wells and then incubated at 37^{o}C with 5% CO_2. After 24 hours, Opti-MEM was removed and replaced with complete media without Geneticin. Cells were incubated for a further 24 hours. For luciferase assays, media was gently removed from the plates by pipetting and cells washed once with DPBS. All excess liquid was removed and 50 µl of a mixture of equal volumes of BriteLite Plus Reagent (Perkin Elmer, Waltham, MA) and DPBS (Ca^+/Mg^+) (Thermo Fisher Scientific) added to each well. After 5 minutes room temperature incubation, bioluminescence was read on the Glomax Multi Detection system (Promega, Southampton, UK) using the Bright-Glo pre-programmed protocol with default setting. Knockdown percent was calculated by subtracting the blank reading and calculating each bioluminescence reading as a percent of the mean of the NSO control. Values were subtracted from 100% to determine the percentage knockdown. EC_{50} and R^2 goodness of fit was compared for each siRNA vs shRNA HCV RNAi analogue by calculating the nonlinear regression curve fit as log (agonist) vs response (three parameters) with the bottom of the curve constrained to zero using GraphPad Prism 7 for Mac OSX.

2.6.2 Knockdown of HCV replicon for collection of RNA for RACE-Seq

For knockdown assays, RNAi oligonucleotides were calculated to give 2X the desired active concentration in the final required volume of Opti-MEM. See table Table 4.1 for final concentrations of oligonucleotides. For example, for siRNA22, the desired concentration was 0.5 nM. Therefore a 1 nM concentration of siRNA22 was prepared in 3000 µl of Opti-MEM by adding together 6 µl of siRNA22 at 500 nM, 59.4 nM NSO at 5 µM and 2,934 µl Opti-MEM. Then 7 µl of Dharmafect3 was added to give ~0.0025 µl/µl Dharmafect3 and the oligo preparation mixed and 50 µl volumes aliquoted to assay plates. The NSO, mock and non-transfection controls were prepared as previously (Section 2.4). After 30-45 minutes incubation to allow complexes to form, 7000 cells in 50 µl was added to each well. As with the dose-response assays, media was changed after 24 hours. Cells were collected after a further 24 hours incubation. Knockdown was determined by a luciferase assay and viability determined by the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenytetrazolium bromide (MTT) assay.

2.6.3 Knockdown of Transthyretin for collection of RNA for RACE-Seq

Annealing of siRNA duplex

To prepare siRNA-TTR, each of the single RNA strands, siRNA-TTR-AS and siRNA-TTR-SS were resuspended to 100 µM in duplex buffer (IDT). To anneal the siRNA-TTR strands, 50 µl of each RNA strand (AS and SS) was pooled in a PCR tube, mixed and heated to 92^{o}C for 2 minutes, then allowed to slowly cool at room temperature.

Reverse Transfection

For knockdown assays, the DsiRNA-TTR, siRNA-TTR and a Scrambled Negative Control DsiRNAscr (1 nmol) (catalogue number 51-01-19-08) (IDT) was prepared at 10 nM by adding together; 52 µl RNAi Oligo at 500 nM, 20.8 µl NSO at 5 µM and 1,227 µl Opti-MEM. 1 µl of Dharmafect3 was added to each preparation, mixed well and 250 µl transferred to four wells of a 24 well plate. NSO control was prepared by mixing together 26 µl NSO (5 µM), 1,274 µl Opti-MEM and 1 µl Dharmafect3 and transferring 250 µl to assay wells. The mock control was prepared as 1 µl Dharmafect3 in 1300 µl Opti-MEM. The non-transfected control was prepared by adding 250 µl Opti-MEM to assay wells. Huh7/con1b cells were prepared by washing two times with warmed DPBS, detached with 2 ml TrypLE Express, washed and cells pelleted by centrifugation at 400 rpm for 5 minutes. Cells were resuspended in 5 ml Opti-MEM. After duplicate cell count, cells were diluted to 168,000 cells/ml and 250 µl cell suspension added to each well to give ~42,000 cells/well. Media was changed after 24 hours to standard DMEM media without G418. After a further 24 hours incubation at 37^{o}C, 5% CO_2, total RNA was either extracted directly from the plate by the TRIzol® method, or cells harvested and stored at -80^{o}C. Cell viability was determined by the MTT assay using the remaining treatment well as described in Section 2.7.1 with the following amendments: 500 µl of prepared MTT reagent (0.05 mg/ml) was added to the 24 well plate. Plates were incubated at 37^{o}C, 5% CO_2 for 4 hours and then the supernatant removed. The Formazan crystals were dissolved by adding 200 µl DMSO to each treatment well. Then 50 µl per treatment was transferred to a 96 well plate and absorbance read at 540 nm as described in section 2.7.1. Routine Cell Assays

2.6.4 Cell viability assay

To assess possible cytotoxic effects of experimental procedures, the MTT reduction assay (Riss et al., 2004) was performed as a measure of cell viability. These assays were carried out in clear 96-well cell culture plates. Media was removed and 100 µl prepared MTT reagent (0.05 mg/mL in DMEM) (Thermo Fisher Scientific) added to each well. Plates were incubated at 37^{o}C in a humidified 5% CO_2 incubator for 4 hours. After incubation, the supernatant was carefully removed and plates either stored at -20^{o}C or processed immediately. Formazan crystals were dissolved by the addition of 50 µl Dimethyl Sulfoxide (DMSO) to each well and plates incubated in the dark for 10 minutes. The blank reagent control consisted of DMSO only. Absorbance was read at 540 nm on the SPECTROstar Nano (BMG LABTECH, Aylesbury, UK) and analysed using the MARS data analysis software. Viability was calculated by subtracting the blank OD value from all reading, then calculating each reading as a percent of the mean NSO control value.

2.6.5 Luciferase reporter assay

Luciferase reporter assays are commonly used to monitor cellular events coupled to gene expression. These reporters are very sensitive since no background luminescence exists in host cells. The HCV replicon in the Con1B cell line contains a firefly luciferase gene *in lieu* of the viral packaging genes (Krönke et al., 2004), producing bioluminescent luciferase protein at amounts proportional to the viral content in the cells. Thus, quantification of bioluminescence in these cells is an appropriate measure of replicon levels (Krieger et al., 2001). To determine the lower limit of detection for luciferase-expressing Con1B cells, the luciferase activity of serially diluted cells was measured. To do this, cells from a single 75 cm^2 flask at 90% confluency were re-suspended in 2 ml DMEM. Duplicate cells counts were determined using the Trypan Blue assay. Cells were diluted at half-logarithmic (3.16 fold) dilution in triplicate in white 96 well plates. To do this, 48,000 cells in 146 µl volume were added to well A1, A2 and A3. Cells were mixed well and 46 µl transferred to the next well that contained 100 µl of DMEM. This was repeated for consecutive wells on the plate. The final cell count was therefore approximately 30,000 cells, 10,000 cells, 3,000 cells, 1,000 cells, 300 cells, 100 cells and 30 cells. Cells were pelleted by centrifuging at 1000 rpm for 5 minutes and media removed by gentle pipetting. For the luciferase assay, an equal volume of DPBS (Ca^+/Mg^+) (Thermo Fisher Scientific) and BrightLite Plus luciferase reagent (PerkinElmer, Coventry, UK) was mixed together and 50 µl added to each well. Blank reagent control consisted of luciferase reagent mix without cells. Bioluminescence was determined on the Glomax Multi Detection system (Promega, Southampton) using the Bright-Glo pre-programmed protocol with default setting. Standard curves were generated by subtracting the blank from readings and then plotting the mean and error vs number of cells per well.

2.6.6 Generating cDNA templates for qPCR

The following method was used to generate the cDNA templates for both the qPCR amplification efficiency experiments as well the qPCR assays for determining *TTR* knockdown. Total RNA was quantified using the Qubit RNA High Sensitivity kit (Thermo Fisher Scientific) and then diluted in nuclease-free water to give 500 ng RNA in 10.5 µl volume. SuperScript IV (SSIV) Reverse Transcriptase (Thermo Fisher Scientific) was used for all reverse transcription assays following the manufactures instructions. Reverse transcription was carried out in 20 µl reactions and prepared by placing the following together in a PCR tube; 10.5 µl total RNA at 500 ng, 1 µl dNTP mix (10 mM), 0.5µl oligo $d(T)_{22}$ (100 µM), 1µl TTR-rt primer at 2 µM. Components were gently mixed by pipetting and then heated at 65°C for 5 minutes in the TC-PLUS thermal cycler (Techne). Samples were cooled directly on ice, then spun down and the reverse transcription mix added. Reverse transcription reaction mix was prepared as a master mix consisting of; 4 µl 5 X SSIV buffer, 1 µl DTT (100 mM), 0.2 µl RNase Inhibitor (40,000 U/ml) (NEB), 1 µl SSIV Reverse Transcriptase enzyme (200

U/ µl) and 1.3 µl water. Then 7 µl was added to each RNA mix, mixed gently, spun down and cDNA synthesis reactions carried out at 55°C for 10 minutes with heat inactivation of Reverse Transcriptase enzyme at 75°C for 10 minutes. cDNA was stored at -20°C and kept on ice when in use.

2.6.7 Validation of primer amplification efficiency

All qPCR reactions were carried out on the Applied Biosystems 7500 Fast PCR Instrument and using Luna Universal qPCR Master Mix (NEB). Primer amplification efficiency for TTR and GAPDH qPCR primers was verified prior to evaluating relative *TTR* mRNA levels after treatment. PCR amplification efficiency was established using calibration curves. To generate the calibration curves, total RNA was extracted from Huh7/con1b cells, DNase I treated, purified and quantified in duplicate using Qubit RNA High Sensitivity kit. Then 500 ng of purified RNA was reverse transcribed using SSIV reverse transcriptase and a pooled primer mix of TTR reverse primer and Oligo $d(T)_{22}$ primer. Primer efficiency was measured using serially diluted cDNA (50 ng, 10 ng, 5 ng, 1 ng, 0.5 ng 0.1 ng and 0.05 ng) as template in qPCR reactions. Reactions were carried out in triplicate and C_T values used to calculate PCR efficiency by plotting average C_T value vs log of cDNA input amount. To test for DNA contamination in RNA extractions, a sample of purified RNA was used as template in the DNA contamination control reactions. Cycle conditions were; 95°C for 30 seconds, followed by 40 cycles of 95°C for 5 seconds, 55°C for 30 seconds, 72°C for 30 seconds (Hayashi et al., 2012). The R^2 and the equation of each standard curve was calculated and the slope value of each standard curve was used to compute primer efficiency percentage and the amplification factor for each primer set using the online qPCR Efficiency Calculator tool from Thermo Fisher Scientific (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library /thermo -scientific-web-tools/qpcr-efficiency-calculator.html).

2.6.8 Relative gene expression by the $2^{-\Delta\Delta CT}$ Method

For determining the relative mRNA knockdown, qPCR reactions were performed in triplicate and included both a non-template control and DNA contamination control (consisting of RNA as template). All qPCR reactions were carried out as 12 µl reactions. Template cDNA for each treatment was generated as described in Section 2.6.6, and diluted 1:2 resulting in approximately 25 ng/µl cDNA template. Then, 2 µl of the diluted template was used for each qPCR reaction. A master mix of reagents was prepared which consisted of 6 µl Luna Universal qPCR Master Mix (NEB), 0.6 µl forward primer, 0.6 µl reverse primer and 2.8 µl water per well. GAPDH was used as internal reference gene. Cycle conditions were as per section 2.6.7. The $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak, 2008) was used to calculate the relative change in gene expression level by real-time quantitative PCR.

2.7 RNA extraction and purification

2.7.1 Collection of cells from 96 well plates

For collection of cells from 96-well plates, cells were washed twice with pre-warmed DPBS and incubated with TrypLE Express reagent at 50 µl per well for up to 8 minutes at 37°C with 5% CO_2. Then, 100 µl of media was added to row one of the plate and cells were mixed at least 7 times to dislodge the cells and then cells transferred to the next column. This was repeated for the whole plate and dislodged cells collected in a 15 ml conical tube. Plates were observed under the microscope to ensure cells had been dislodged and collected. Cells were centrifuged at 1000 rpm for 5 minutes at 4°C. Supernatant was removed and cell pellets stored at -80°C for RNA extraction. Cells were collected from 36 wells of a 96-well plate for RACE-Seq assays.

2.7.2 RNA extraction using the PureLinK RNA Extraction Kit

Total RNA was isolated using the PureLink RNA Mini Kit (Life Technologies, UK). Cell pellets were removed from -80°C storage and 600 µl lysis buffer containing 1% v/v 2-mercaptoethanol added and vortexed for a full 1 minute. After addition of equal volume of 70% v/v ethanol and a brief vortex, samples were transferred to spin columns. Columns were centrifuged at 12 000 x g for 15 seconds, then washed with wash buffer I and wash buffer II. Total RNA was eluted in 50 µl RNase free water. RNA was stored in two aliquots at -80°C.

2.7.3 Extraction of RNA by TRIzol® method

Total RNA was extracted directly from 24 well plates or pelleted cells using TRIzol® reagent (Thermo Fisher Scientific). For the extraction of RNA directly from 24 well plates, cells were washed twice with approximately 1 ml warmed DPBS and all excess liquid removed. For each knockdown treatment, the extract from three wells was pooled by adding 200 µl of TRIzol® to each well, mixing vigorously and then transferring the lysate to a 1.5 ml microcentrifuge tube containing 1.5 µl GlycoBlue^{TM} Coprecipitant. Chloroform at 200 µl/1ml TRIzol$^{®}$ was added to each tube and contents mixed by inversion for 10 seconds then incubated at room temperature for 5 minutes. After centrifugation at 12,000 x g for 15 minutes at 4^{o}C, the upper phase was carefully transferred to a new tube and isopropanol at half the original volume of TRIzol$^{®}$ added. Samples were gently mixed by inverting three times and the samples incubated at room temperature for 10 minutes. After centrifugation at 12,000 x g for 10 minutes at 4^{o}C, the supernatant was gently removed leaving behind a blue pellet. Pellets were washed with 500 µl 75% ethanol by gentle inversion and flicking, centrifuged at 10,000 x g for 5 minutes at 4^{o}C and the ethanol removed. Pellets were allowed to air dry and the purified RNA resuspended in 44 µl nuclease-free water.

Deoxyribonuclease I Treatment

RNA was DNase treated by adding 5 µl of 10x DNase buffer (Thermo Fisher Scientific) to each 44 µl volume of total RNA, then 1 µl of Deoxyribonuclease I (DNase I) (1U/ml) (Thermo Fisher Scientific) was added, samples mixed and incubated at 37^{o}C for 10 minutes on a heat block. To deactivate DNase I, ethylendiaminetetraacetic acid (EDTA) was added to a final concentration of 1 mM, mixed and samples heated to 85^{o}C for 10 minutes.

Final purification

DNase treated RNA was purified by Phenol-chloroform extraction. Reactions were made up to 100 µl with nuclease-free water and an equal volume of phenol/chloroform added, and centrifuged at 12,000 x g for 5 minutes. The aqueous phase was removed to a new tube containing 1.5 µl of 15 mg/ml GlycoBlue^{TM} Coprecipitant (Thermo Fisher Scientific), 10 µl 3M Na acetate, pH 5.2 (Thermo Fisher Scientific) and 200 µl (2x volume) of absolute ethanol (VWR). Samples were vortexed briefly and incubated for 30 minutes on dry ice. Samples were then centrifuged at 16,000 x g for 20 minutes at 4^{o}C and the supernatant carefully removed. The RNA pellet was washed once with 500 µl 70% v/v ethanol, with mixed by inversion, then vortexed for 10 seconds. After centrifugation at 16,000 x g for 2 minutes the supernatant was carefully removed and RNA pellets were air dried for up to 5 minutes, resuspended in 30 µl RNase-free water and kept on ice.

RNA analysis and quantification

RNA quality was assessed by Nanodrop and RNA quantified using the Qubit RNA High Sensitivity kit.

2.8 Assessment of nucleic acids

2.8.1 NanoDrop quantification (DNA and RNA)

Concentration and purity of nucleic acid extractions were determined using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). 1 µl of purified nucleic acid was analyzed using default setting for DNA or RNA as required.

2.8.2 Qubit quantification

DNA or RNA quantification was assessed using either the Qubit dsDNA HS assay kit (Thermo Fisher Scientific) or the Qubit RNA HS assay kit (Thermo Fisher Scientific) following the manufacturer's protocol. Briefly, Qubit working solution was prepared by diluting Qubit reagent 1:200 with the Qubit buffer. 1 µl of sample was then added to 199 µl of working solution in Qubit assay tubes. Standards were prepared by diluting 10 µl of each standard in 190 µl of working solution in Qubit assay tubes. After 3 minutes incubation, standards and samples were read on the Qubit 2.0 Fluorometer.

2.9 RNA Ligase-Mediated Rapid Amplification of cDNA Ends (RLM RACE)

An outline of the RLM RACE assay is illustrated in Figure 1.6. Adapter ligation reactions, were prepared as either 10 µl volumes, or scaled to 20 µl volumes. For the 20 µl reactions, 2 µl of truncated GeneRacer RNA adapter (100 µM) and >1 µg of total RNA (total volume 12.5 µl) was heated at 65^{o}C for 5 minutes in a thermal cycler then snap cooled on ice for 2 minutes. Then, 5.5 µl of a ligation mix (consisting of 2 µl of 2x ligase buffer (NEB, Hitchin, UK), 2 µl 10 mM Adenosine 5'-Triphosphate (ATP) (NEB), 0.5 µl RNase Inhibitor, Murine (40,000 units/ml) (NEB) and 1 µl T4 RNA ligase (10,000 units/ml, NEB) was mixed by pipetting and briefly centrifuged at 10 000 x g. Adapter ligation was carried out at 37^{o}C for 1 hour in a Techne TC-PLUS thermal cycler (Bibby Scientific, Staffordshire, UK) with heated lid setting turned off. RNA precipitation was carried out by bringing the reaction volume to 100 µl with nuclease-free water, then 100 µl of acid phenol/chloroform (Thermo Fisher Scientific) was added to each adapter ligation reaction and vortexed for 30 seconds. PCR tubes were centrifuged at 12 000 x g for 5 minutes at room temperature and approximately

100 µl of the upper phase was carefully removed to a 1.5 ml nuclease-free centrifuge tube. 1.5 µl of 15 mg/ml GlycoBlue (Thermo Fisher Scientific) and 10 µl 3M sodium (Na) acetate, pH 5.2 (NEB) and 200 µl (2x volume) of absolute ethanol (VWR) was added to each volume and vortexed briefly. After 30 minutes incubation on dry ice to precipitate the RNA, tubes were centrifuged at 16 000 x g for 20 minutes at 4^{o}C. Being careful not to disturb the pellet, tubes were carefully removed from the centrifuge and the supernatant removed by pipetting. The pellet was washed by adding 500 µl of 70% v/v ethanol and mixed by inversion, then vortexed for 10 seconds. After centrifugation at 16 000 x g for 2 minutes, supernatant was carefully removed. Tubes were centrifuged for a further 10 seconds at 16 000 x g and residual ethanol removed. RNA pellets were air dried for up to 5 minutes and RNA re-suspended in 11 µl RNase-free water and kept on ice.

First strand cDNA was synthesised using either SuperScript III reverse transcriptase enzyme (Thermo Fisher Scientific) or Moloney Murine Leukemia Virus (M-MuLV) Reverse Transcriptase (NEB). First, 11 µl of RNA, 1 µl of the gene-specific reverse primer at 2 µM and 1 µl of dNTP (10 mM each, Sigma-Aldrich, Poole, UK) were mixed together and then heated for 5 minutes at 65^{o}C to remove secondary structure and samples snap cooled on ice for 2 minutes.

For SuperScript III: A master mix was prepared consisting of: 4 µl 5X First strand buffer, 1 µl of 0.1 mM DTT, 0.2 µl RNase Inhibitor, 1 µl of SuperScript III reverse transcriptase enzyme and the mix made up to 7 µl volume using nuclease-free water was added to the RNA.

For M-MuLV: A master mix was prepared consisting of: 2 µl 10X reaction buffer, 0.2 µl RNase Inhibitor, 1 µl of M-MuLV reverse transcriptase enzyme and the mix made up to 7 µl volume using nuclease-free water. This was added to the prepared RNA. Reverse transcription was carried out at 55^{o}C for 30 minutes and enzyme inactivated by increasing the temperature to 70^{o}C for 15 minutes. cDNA was stored at -20^{o}C or used immediately in PCR reactions.

2.9.1 PCR amplification

The cDNA was amplified by anchored PCR reaction using forward primer, GreneRacerF1 and a gene-specific reverse primer. PCR reactions were set up as 40 µl volume reactions as; 20 µl Q5 Hot Start High-fidelity 2X Master Mix (NEB), 2 µl GeneRacerF1 primer (10 µM), 2 µl gene-specific reverse primer (10 µM), 4 µl cDNA template (10 % of PCR volume) and made up to 40 µl with nuclease-free water (Thermo Fisher Scientific). PCR amplification was carried out on the Techne TC-PLUS thermal cycler (Bibby Scientific) with the following cycle conditions; 98^{o}C for 30 seconds, 35 cycles of (98^{o}C for 15 seconds, 60^{o}C for 10 seconds, 72^{o}C for 15 seconds), then 72^{o}C for 2 minutes and hold at 10^{o}C.

When second-round PCR was required, first round PCR was scaled down to 15 µl reactions and PCR amplification carried out as the above conditions. Then, residual primers were enzymatically degraded and dNTPs dephosphorylated using the NEB Rapid PCR Clean up Enzyme Set (NEB). To a PCR tube, 5 µl of first-round PCR reaction was added to 1 µl of Exonuclease I (Exo I) enzyme and 1 µl of Shrimp Alkaline Phosphatase (rSAP) enzyme. The contents were mixed gently, briefly centrifuged and placed in the thermal cycler at 37^{o}C for 5 minutes followed by 80^{o}C for 10 minutes. Then, 4 µl of cleaned PCR was used in a 40 µl PCR reaction as above using GeneRacerF2 primer and a gene specific primer located at a position internal to the first PCR primer sequence. Up to 10 µl of PCR reaction was analysed by electrophoresis on a 2 %TBE agarose gel. The remaining PCR reaction was used for library preparation

2.9.2 Agarose gel electrophoresis

Gels of 2% w/v agarose was prepared by dissolving agarose (Appleton Woods, Birmingham, UK) in 50 ml 1 X Tris/Borate/EDTA (TBE) buffer (Thermo Fisher Scientific) in a microwave until all solid particles were completely dissolved. The heated solution was allowed to cool for 10 minutes and 2 µl of SYBR Safe DNA staining dye (Thermo Fisher Scientific) was added to the gel. The agarose was poured into its gel tank and allowed to solidify. Then, 5 - 20 µl of PCR sample was diluted 1:4 with 6X Orange DNA loading Dye (Thermo Fisher Scientific) before being loaded onto the gel. 5 µl of pre-dyed ladder, GenerRuler 100 bp DNA ladder (Thermo Fisher Scientific) was loaded per well. Gels were run at 100 volts for 40 - 60 minutes prior to UV analysis.

2.9.3 Gel extraction

Bands were visualised under blue light on the Safe Imager 2.0 Blue Light Transilluminator (Thermo Fisher Scientific) and target bands excised from the gel using a scalpel. DNA was extracted from gel slices using the NEB Monarch DNA Gel Extraction Kit (NEB). Briefly, each gel slice was dissolved in 4 volumes of gel dissolving buffer at 50^{o}C for 5-10 minutes and loaded to the Monarch DNA clean up column. Columns were centrifuged at 13 000 rpm for 1 minute followed by two washes with DNA Wash Buffer. Purified DNA was eluted in 10-20 µl DNA elution buffer (NEB) or nuclease-free water.

2.10 Ion PGM library preparation

2.10.1 AMPure bead clean (amplicons >100 bp)

The Agencourt AMPure XP bead system uses solid-phase paramagnetic beads in an optimised polyethylene and salts buffer. DNA binds to the beads in a concentration-independent manner, but associates with the beads in a size-dependent manner when adjusting the volume of bead suspension ratio to DNA solution. Bound DNA is separated from the supernatant and excess primers, nucleotides and salts using a magnet. The separated DNA is washed and then purified DNA eluted from the beads (Rodrigue et al., 2010). Amplified libraries were bead cleaned by adding 1.6X volumes of room temperature Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) and mixed thoroughly by pipetting. Samples were then incubated on the bench for 5-8 minutes before being placed on a magnetic rack for 3 minutes. The supernatant was removed and beads were washed twice with 300 µl of 70% v/v ethanol. Residual ethanol was removed and beads were air dried for 3-5 minutes. Purified DNA was re-suspended in 25 µl Tris/EDTA (TE) (supplied with library preparation kit), vortexed briefly, centrifuged at 0.3 x g for 5 seconds and placed on the magnetic rack for 2 minutes. Purified DNA was collected in the supernatant (~23 µl) and used for library preparation.

2.10.2 AxyPrep Mag PCR clean (amplicons <100 bp)

A mix of 70 µl of isopropanol and 180 µl of AxyPrep beads was prepared. A 25 µl volume of this mix was added to each 10 µl volume of PCR product, mixed by pipetting, and incubated for 5 minutes at room temperature. Tubes were then transferred to a magnetic rack for 3 minutes, the cleared supernatant removed and magnetic bead pellets washed two times with 200 µl 70% v/v ethanol without disturbing the pellet. Residual ethanol was removed by centrifugation at 0.3 x g for 5 seconds. Beads were allowed to air dry for up to 5 minutes and DNA re-suspended in 40 µl TE buffer.

2.10.3 5' RACE library preparation using NEBNext Library Prep Set

Amplicon libraries were generated using the NEBNext Fast DNA Prep set for Ion Torrent (NEB) as outlined in Figure 3.8. First, amplicons were bead cleaned using either the AxyPrep Mag PCR beads for recovery of fragments less than 100 bp or the Agencourt AMPure XP beads for PCR amplicons greater than 100 bp. Purified DNA was quantified by Qubit HS DNA assay kit and then 200-500 ng DNA was used in the end repair reaction, which consisted of 17 µl purified DNA, 2 µl NEBNext End repair reaction buffer and 1 µl NEBNext End repair enzyme mix. DNA was blunt-ended in the end repair reaction with a 20 minute incubation in a thermal cycler set to 25^{o}C followed by 10 minutes at 70^{o}C. The adapter ligation mix was prepared as a master mix consisting of: 6 µl nuclease free water, 3.3 µl T4 DNA ligase buffer for Ion Torrent, 0.9 µl P1 adapter, 0.9 µl Ion Xpress Barcode Adapter (Thermo Fisher Scientific), 0.3 µl BSt 2.0 WarmStart DNA Polymerase and 2 µl T4 DNA ligase. This (13.4 µl) was added to the end repair reaction, mixed and placed in thermal cycler at 25^{o}C for 15 minutes followed by 5 minutes at 65^{o}C. Barcoded reactions were bead cleaned using Agencourt AMPure XP beads and purified DNA eluted in 30 µL TE buffer. Samples were then amplified using NEBNext Q5 Hotstart HiFi PCR Master Mix and the sequencing amplification primer set supplied in the kit. Samples were amplified in 7 cycles of PCR with initial denaturation 98^{o}C for 30 seconds, then 7 cycles of (98^{o}C for 10 seconds, 58^{o}C for 30 seconds, 65^{o}C for 30 seconds) followed by 5 minutes at 65^{o}C and hold at 10^{o}C.

2.10.4 Library size selection using the Labchip XT instrument

Labchip XT 300 bp kit (Perkin Elmer) contained all the reagents for DNA extraction. The chip was prepared as per manufacturer instructions and wells to be loaded washed with chip buffer. LabChip XT Software version 2.1.1233.0SP1 (Perkin Elmer) was launched and a run set up. Size range was specified as the expected size of 5' RACE libraries ± 10% and the extract and pause operation selected. Samples were prepared by mixing together 10 µl of DNA and 2 µl of the 6X sample buffer. Samples were vortexed briefly, spun down and kept on ice. The DNA marker was prepared by adding 2 µl of ladder to 2 µl 6X sample buffer and bringing the volume up to 12 µl with nuclease-free water or TE buffer.

Then 15 µl of XT DNA dye was added to the waste reservoir well and mixed by gently rocking the chip. Then, 20 µl of collection buffer was added to collection wells and 20 µl of stacking buffer was carefully added to sample wells and any bubbles removed. The entire 12 µl volume of sample or DNA marker was loaded to sample wells by inserting the tip of the pipette to the bottom of the well and slowly but smoothly pipetting the dye labelled sample beneath the stacking buffer. The chip was run till completion and size selected fragments collected from the collection well and kept on ice.

2.10.5 Library size selection using the Size Select 2% E-gel system

The Size Select™ E-gel® (Thermo Fisher Scientific) is a double-comb, precast agarose gel system that allows PCR products to be collected directly from the gel, without the need to cut sections from the gel and column purify. The gels contain a proprietary fluorescent nucleic acid stain, allowing migrating DNA bands to be visualised in real time on the E-gel iBase™ blue light transilluminator (excitation/emission 490/522 nm). Samples were prepared by diluting 1:5 with 2X Orange DNA loading Dye and added to each well of the Size Select™ 2% E-gel® and 10 µl of 25 bp DNA ladder (Thermo Fisher Scientific) loaded to the central well. All well volumes were made up to 25 µl using nuclease-free water. The gel was set to run the pre-set program (number 8). Migration of DNA bands was monitored until the band of interest reached the reference line. The run was paused and collection wells topped up to 25 µl with nuclease-free water. The run was re-started and bands were monitored as they flowed into the collection well where the sample was pipetted out of the well and transferred to 1.5 ml low-bind Eppendorf tubes (VWR, UK) where they were kept on ice or stored at -20°C.

2.10.6 Gel extraction and purification of NGS libraries

RACE-Seq libraries were excised from 2% agarose gels and DNA purified as in section 2.10.3.

2.10.7 DNA fragment analysis

The Bioanalyzer system (Agilent, Santa Clara, CA) and accompanying High Sensitivity DNA Kit were used to assess the NGS libraries. The Agilent Bioanalyzer 2100 system uses micro-capillary electrophoresis on a chip to rapidly analyse DNA fragments with high sensitivity. The gel-dye mix was first prepared by adding 15 µl of the blue dye concentrate to the DNA gel matrix, mixing by vortexing, centrifuging to pellet the contents and the entire volume was transferred to a spin filter tube. The mix was centrifuged at 2,300 x g for 10 minutes and kept at 4^{o}C for up to six weeks. On the day of assay, all reagents were allowed to equilibrate to room temperature for at least 30 minutes. Samples were prepared by first pipetting 9 µl of the gel-dye mixture into the appropriate well and forcing the mixture into the micro-channels by applying pressure to the well via a 1 mL syringe on the chip priming station. Then 9µl of the gel-dye mixture was also added to the other two wells for gel. All the ladder and sample wells were loaded with 5 µl of the DNA marker before loading of 1 µl of either the molecular size ladder or 1 µl of sample to wells on the chip. After mixing by vortexing at 2400 rpm on the IKA Vortex Mixer, (Agilent) the chip was immediately inserted into the Bioanalyzer 2100 instrument (Agilent Technologies), the appropriate DNA assay selected and the programme started. On completion of the run, the Agilent 2100 Bioanalyzer software identifies DNA peaks between the lower and upper marker peaks, reporting size range and fragment length as well as picomolar quantification of fragment range. After quantification on the Bioanalyzer, samples were each diluted to 1000 pM and pooled in equimolar concentration by adding the same volume of each sample to a low-bind Eppendorf. Pooled libraries were again quantified in duplicate on a Bioanalyzer High Sensitivity DNA chip prior to dilution to 13 to 20 pM for input to emulsion PCR.

2.11 Preparation and sequencing of NGS samples

2.11.1 Emulsion PCR and enrichment

DNA fragments are templated to proprietary sequencing beads, the Ion Sphere Particles (ISPs) using the Ion Torrent OneTouch2 (OT2) instrument (Thermo Fisher Scientific) and accompanying kit Ion PGM Hi-Q OT2 kit (Thermo Fisher Scientific). The Ion OT2 instrument was first cleaned by running the cleaning protocol utilising a used amplification plate and tubes on the instrument. The used consumables are removed and new collection tubes containing 150 µl breaking fluid inserted to the centrifuge. A new plate was inserted and the tubing and needle put in position. The level of the Oil Ion OneTouch Reagent tube was filled to half and the Recovery Solution Ion OneTouch Reagent tube was filled to one-quarter level. The emulsion PCR reactions were prepared by adding in order, to a 2-ml Ion PGM Reagent Mix tube containing 800 µl of reagent mix, 25 µl nuclease-free water, 50 µl Ion PGM Enzyme Mix, 25 µl diluted library and 100 µl Ion PGM Hi-Q ISPs. After vortexing for 5 seconds, the mix was transferred to an Ion OneTouch Reaction Filter as a single 1000 µl volume by slowly pipetting through the sample port. The reaction tube was placed on the OT2 instrument and the 200 bp programme executed. After the run, the ISP pellet was collected by removing all but 50 µl of supernatant from the collection tubes. Then 500 µl of Ion Wash Solution (Thermo Fisher Scientific) was added to each tube, mixed well and transferred to a 1.5 ml low-bind Eppendorf tube. The ISPs are pelleted by centrifugation at 15,500 x g for 3 minutes and all but 100 µl supernatant removed. A 2 µl volume was transferred to a PCR tube for Quality assessment, and the remainder was placed into well 1 of an eight-well strip. The supernatant from 13 µl of Dynabeads™ MyOne™ Streptavidin C1 Beads was removed by placing the beads in a 1.5 ml Eppendorf tube on a magnetic rack. Then 130 µl of MyOne™ Beads Wash Solution was added and the beads re-suspended. The beads were added to well 2 of the 8-well strip. Melt-Off Solution was prepared by mixing together 280 µl Tween™ Solution and 40 µl 1 M NaOH, and transferring 100 µl to well 7 of the 8-well strip. 300 µl of Ion OneTouch™ Wash Solution was added to wells 3, 4 and 5. A new tip was installed in the Tip Arm and 10 µl of Neutralization Solution added to a 0.2 µl PCR tube in the collection well. The programme was initialized. Immediately at the end of the run, the collection tube was mixed by inversing five times. The enriched ISPs were kept at 4^{o}C for up to 3 days.

2.11.2 Quality assessment of templated ISP

The 2 µl sample of ISPs was quality assessed for enriched ISPs using the Ion Torrent Quality Control (QC) kit. Briefly, 19 µl of annealing buffer was mixed with 1 µl of Probe in a PCR tube containing 2 µl of ISP sample. The probe mix was incubated in a thermal cycler at 95^{o}C for 2 minutes then 37^{o}C for 2 minutes and immediately placed on ice for 2 minutes. Excess probe was washed by addition of 200 µl of Quality Control Wash buffer, vortexing and centrifugation at 15 500 x g for 2 minutes. All but 10 µl was removed, without disturbing the pellet and a second wash performed. Then the ISPs were resuspended to a total volume of 200µl in Quality Control Wash buffer which was transferred to a Qubit tube. Negative control consisted of 200 µl Quality Control Wash Buffer. Samples were read on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific) and QC percent calculated with a QC pass accepted at 10 to 30% as recommended by the manufacturer. Samples passing these criteria were progressed to sequencing on the Ion Torrent PGM.

2.11.3 Preparation of sequencing reaction and chip loading

On the day of sequencing, a chlorite wash and water wash was performed. The reagent bottles were washed and prepared as per manufacturer protocols and placed on the instrument and the instrument initialized. At this stage, baseline pH is set by mixing the solutions. Once the initialization had passed the required criteria, the dNTPs were added to the instrument and the new sequencing chip checked. The samples were prepared for sequencing by first adding 5 µl of Control ISPs (Thermo Fisher Scientific) to the enriched ISPs and then centrifuging at 15,000 x g for 2 minutes. All but 15 µl of the supernatant was removed and 12 µl of sequencing primer added to give a total volume of 27 µl. After mixing and centrifuging to collect the contents, the sequencing primer was annealed to the ISPs by heating for 2 minutes at 95^{o}C followed by 2 minutes at 37^{o}C and then kept at 25^{o}C. The checked chip was removed from the instrument and all liquid removed from the from the chip using a pipette. The sample was removed from the thermal cycler and 3.3 µl of Ion PGM Sequencing Hi-Q polymerase added, mixed thoroughly and then 30 µl taken up into the pipette. The chip was slowly filled by winding down the pipette at 0.5 µl per second. After first round centrifugation in the MiniFuge (Thermo Fisher Scientific), the sample was mixed on the chip by tilting the chip to a 45^{o} angle and slowly mixing 25 µl out and back into the chip three times without creating bubbles. The chip was centrifuged again on the MiniFuge for 30 seconds. All liquid was removed from the chip, and the chip transferred to the Ion PGM. The prepared sequencing templates was selected on the PGM and run option selected. The sequencing run was monitored on the Ion Server. At the end of the run, the reagent bottles were removed and the Instrument washed. The sequencing data becomes available on the Ion Torrent Server on completion of the run.

2.12 RACE-Seq data analysis

All bioinformatics analysis for this project was carried out on an Apple MacBook Pro, 8 GB memory, macOS Sierra version 10.12.5. The following command line packages were used: *cutadapt* (version 1.8.3) (http://cutadapt.readthedocs.io/en/stable/index.html) and FASTX

ToolKit (version 0.0.13) (http://hannonlab.cshl.edu/fastx_toolkit/). The alignment and 5' end counting was performed using adapted versions of the RACE-SEQ-lite pipeline (Theotokis et al., 2017). The pipeline performs the alignment of RACE-Seq data to the reference using Bowtie and the 5' ends counted using SAMtools. The programme is presented as a custom R programme and executed in RStudio. To run the file, the processed RACE-Seq data file as a .FASTQ format file, the RACE-SEQ-lite pipeline file and the reference file as a .fasta format file was placed into the same folder. The RACE-SEQ-lite file was opened in RStudio, the region of the RISC hybridization site adjusted and the mismatch criteria set. The pipeline was then executed to generate a comma-separated value (CSV) file containing each position on the reference sequence, the corresponding nucleotide, the depth of coverage reported as 5' end counts, the percent of aligned reads at each position and the $\log_{10}$ transformation of read counts.

2.12.1 Filtering of RACE-Seq datasets

At the end of the sequencing run, files for each RACE-Seq sample was obtained from the Ion Torrent Server as de-multiplexed FASTQ files. Each file was assigned with its barcode and sample name as assigned prior to sequencing. The Ion Torrent adapter sequences had been auto removed from reads. To prepare samples for alignment to the reference, samples were first pre-filtered to specifically select for high quality target specific reads. Each set of data was filtered by applying the sequence of queries below, which were executed line-by-line in the command line environment (Table 2.4).

The following sequence of actions was used for filtering of reads to generate a 'filtered dataset' with reads >30-nt length (see Table 2.4 for the accompanying code).

I. Reads were filtered by identifying the truncated GeneRacer adapter sequence on the 5' end of reads and the reverse complement of the GeneRacer adapter sequence on the 3' end of reads.

II. Reads containing the reverse complement sequence of the GeneRacer adapter at the 3' end were reverse complemented using the FASTX toolkit.

III. The reverse complemented file was concatenated to the RACE.fastq file.

IV. The GeneRacer adapter sequence was trimmed from all reads with --no-indels, that sets the insertion and deletion tolerance to 0.

V. The sequence of the reverse primer was identified on the 3' end of reads and reads partitioned to a new file with either trimming of the sequence for RACE-Seq long amplicon assay or no-trimming for the legacy RACE-Seq assays.

VI. The 3' ends of reads were quality trimmed.

VII. Short reads were removed (with minimum length set at 30-nt for the legacy RACE-Seq assays.)

The datasets were then analysed using an adapted RACE-SEQ-lite pipeline where the adapter trimming code has been removed since adapter trimming had been done during the filtering process.

For aligning the reads to the reference sequence, the Bowtie aligner can allow up to three mismatches. Mismatch criteria was generally set to zero for aligning RACE-Seq data. However, the quasispecies nature of the HCV RNA genome means that there is naturally a high amount of sequence variability from the amplified target. Therefore, there would always be a subset of data that failed to align and therefore failed to be included in the RACE-Seq analysis. For the RACE-Seq data analysis, the region immediately after the RNA adapter is really what is of primary importance. Therefore, in order to capture more of the RACE-Seq relevant dataset, an alternative filtering protocol was implemented to shorten the reads prior to alignment (Tabernero et al., 2013; Ganesh et al., 2016) and thus eliminate the part of the read that may hold mismatches to the target that exclude the read from alignment. The read quality of the shortened datasets was analysed using the FASTQC tool version 0.11.4.

The following sequence of actions was used for filtering of reads to generate the' filtered dataset' with reads of adapter+20-nt, see Table 2.5.

Steps (i), (ii) and (iii) were carried out as before using the code as per Table 2.4.

iv. The reverse primer sequence was identified on the 3' end of reads and trimmed.

v. For the reads the held the reverse primer trimmed, reads were filtered for those reads that additionally held the sequence of the next 10 bases immediately internal the reverse primer position. This action specifically enriched for the amplified target and eliminated non-specific amplified reads.

vi. Reads were shortened to contain the RNA adapter sequence and the next 20-nt after the adapter using 'fastx_clipper'. The last 4 bases of the RNA adapter sequence were excluded from the sequence at this stage to account for variability at the 5' end of the sequence adapter that may have occurred during the Ion Torrent Server trimming of the sequencing adapter sequences from reads.

This adapter+20-nt dataset constituted the filtered dataset. The RACE-SEQ-lite pipeline trimmed the RNA adapter sequence and performed the read alignment and 5’ end counts.

Table 2.4 Command line code used for filtering RACE-Seq datasets

	Command line code
i	cutadapt -a revRACE=TTTCTACTCCTTCAGTCCATGTCAGTGTCC -g RACE=GGACACTGACATGGACTGAAGGAGTAGAAA --no-trim -o {name}.fastq input_name.fastq
ii	fastx_reverse_complement -Q33 -i revRACE.fastq -o name1.fastq
iii	cat RACE.fastq name1.fastq >name2.fastq
iv	cutadapt -g name3=^GGACACTGACATGGACTGAAGGAGTAGAAA --no-indels -o {name}.fastq name2.fastq
v	cutadapt -a name4=reverse_primer_sequence -o {name}.fastq name3.fastq
v*	cutadapt -a name4=reverse_primer_sequence --no-trim -o {name}.fastq name3.fastq
vii	cutadapt -q 30 -o name5.fastq name4.fastq
viii	cutadapt --minimum-length 30 name5.fastq > final_output_name.fastq

v* = code for selecting reads without trimming the reverse primer sequence

Table 2.5 Command line code for adapter+20-nt filtering of RACE-Seq data

	Command line code
i	cutadapt -a revRACE=TTTCTACTCCTTCAGTCCATGTCAGTGTCC -g RACE=GGACACTGACATGGACTGAAGGAGTAGAAA --no-trim -o {name}.fastq input_name.fastq
ii	fastx_reverse_complement -Q33 -i revRACE.fastq -o name1.fastq
iii	cat RACE.fastq name1.fastq >name2.fastq
v	cutadapt -a name3=reverse_primer_sequence -o {name}.fastq name2.fastq
v	cutadapt -a name4=internal-10-nt -o {name}.fastq name3.fastq
vi	fastx_clipper -Q33 -a ACTGACATGGACTGAAGGAGTAGAAA -d 20 -i name4.fastq -o final_output_name.fastq

2.12.2 Data presentation

Bar chart plots of 5’ end counts vs nucleotide within the RISC hybridization site or count as a percent of total aligned reads were generated.

2.12.3 Generating hybridization images for RNAi bioactives

Hybridization images for the RNAi molecules were generated using Unafold online server (http://unafold.rna.albany.edu/) using the ‘Two-state melting (hybridization)’ programme for siRNA and Dicer substrate RNAi and the ‘Two-state Folding’ programme for shRNA using default settings (Zuker, 2003).

2.12.4 Analysis of unaligned data

Reads that did not align to the HCV replicon reference sequence when mismatch criteria was set to 3, were retained for reanalysis. To facilitate manual review, the reads were collapsed and then the top 100 sequences were aligned to the reference sequence using T-Coffee, (https://www.ebi.ac.uk/Tools/msa/tcoffee/) an online multiple sequence aligner. This aligner uses a combination of local and global alignment and generates a position-specific scoring scheme in a progressive alignment method. Additionally, since values obtained for a given pair of sequences is evaluated against information from the other sequences in the set, a significant increase in alignment accuracy is obtained (Notredame et al., 2000) To better visualise the output, MView (http://www.ebi.ac.uk/Tools/msa/ mview/) was used. This is a tool that converts the results of a multiple sequence alignment to a coloured format where hits are stacked against the query. Pyrimidines (CT) appear as light blue while purines (AG) are dark blue (Brown et al., 1998). Default settings were used.

3 TRANSITIONING RACE-SEQ TO THE ION TORRENT PGM

3.1 Introduction

The 5’ RLM-RACE assay is a semi-specific targeted assay, that is used to capture and enrich for the downstream cleaved product generated by the catalytic activity of AGO2-RISC when directed to a particular site on the target RNA by the guide strand of an siRNA. The downstream cleaved RNA fragment of the targeted mRNA is left with a monophosphate on the newly formed 5’ end (Martinez and Tuschl, 2004; Schwarz et al., 2004). After total RNA extraction, T4 RNA ligase is used to attach the RNA adapter to the 5’ end of the cleaved fragments (Soutschek et al., 2004; Frank-Kamenetsky et al., 2008; Judge et al., 2009). T4 RNA ligase catalyses the formation of a phosphodiester bond between the terminal 5’ phosphate of mRNA and the 3’-hydroxyl-terminated RNA adapter, in a template independent manner. Any RNA or single stranded DNA with a 5’ phosphate will be adapter ligated. The target sequence of interest is enriched by cDNA synthesis with a gene-specific reverse transcription primer and amplification using a primer to the RNA adapter and the mRNA of interest. The PCR products are cloned into a vector and individual clones sequenced by Sanger Sequencing to confirm the expected cleavage site (Davis et al., 2010; Chang et al., 2010; Clark et al., 2013).

At the start of this project in early 2014, only two publications had used NGS of 5’ RLM-RACE products to interpret RISC activity of introduced siRNAs in humans or human cell lines. In both cases Illumina sequencing was used (Tabernero et al., 2013; Denise et al., 2014). In a first-in-humans trial of an RNAi therapeutic targeting vascular endothelial growth factor (*VEGF*)-A mRNA and kinesin spindle protein (*KSP*) mRNA in cancer patients, Tabernero et al., (2013) sequenced the 5’ RACE-PCR products from 15 patients both pre- and post-treatment. The pre-treatment 5’ RACE data was used to set a base level for specific cleavage product. Although the authors did not present all their data, they confirmed that for two patients, for an average of three measurements from the same biopsied tumour sample, the 5’specific cleavage product for *VEGF* mRNA reached >20% above base level and for a further third patient was 4.3% above base level. For *KSP* mRNA, no specific cleavage product was detected by 5’ RACE, and this was attributed to the much lower levels of *KSP* mRNA compared to *VEGF* mRNA. The authors also commented that the mix of tumour tissue types in the biopsied material (normal liver, necrotic/fibrotic tissue and viable tumour tissue) may have impacted mRNA measurement assays.

Denise et al. (2014) sought to use NGS to gain new insights to the processing and activity of the shRNAs expressed from TT-034, a DNA-directed RNAi therapeutic targeting HCV. In an attempt to comprehensively characterise various aspects of drug activity the hairpin sequences expressed from TT-034 after transfection to Huh7/con1b cells were isolated and sequenced by NGS. In analysing these sequences they observed that hairpin expression from TT-034 was uneven due to variation in the expected transcription start site (Lavender et al., 2012). They reported that up to 95 siRNA strands could be processed by Dicer cleavage, with individual siRNA strands having variable sequences at either of their ends and some of the siRNA strands with TT-034 backbone sequences or sequences from the shRNA loop structure. Thus, unexpected siRNAs were generated from TT-034. The most common siRNA sequences were selected and synthetic siRNAs purchased and transfected to Huh7/con1b cells. RACE-Seq analysis was conducted and sequences analysed. They found cleavage products with 5' ends upstream of the expected cleavage site for siRNA19 and siRNA22 which led to the proposal that AGO2-RISC might favour a secondary cleavage position beyond the expected cleavage site (between bases 10-11).

In this part of the project, the legacy RACE-Seq assay reported by Denise et al. (2014) was carried out using a synthetic siRNA22 and synthetic shRNA22. A schematic of the assay is presented in Figure 3.1 and indicates the position of the reverse primer, the PCR amplification primers and the expected size of the 5' RACE-PCR product. In 2014, the Ion Torrent PGM Platform was still new to the market, having been launched in 2010. It offered a cheaper alternative to the Illumina sequencing platform offering broader access to NGS technology, and was already showing promise in clinical diagnostics (Singh et al., 2014; Malapelle et al., 2015).

The main objective of this chapter was therefore to perform the RACE-Seq assay on the Ion Torrent PGM. To do this, total RNA after knockdown of HCV genomic RNA was collected and 5' RML-RACE performed. The amplicons were progressed to NGS library preparation and libraries passing validation progressed to emulsion PCR and sequencing on the Ion

Torrent PGM. The aligned and unaligned RACE-Seq datasets were comprehensively analysed.

The following approaches were taken towards achieving the objectives:

- HCV replicon transcripts were targeted using the synthetic siRNA22 and shRNA22 analogues of TT-034.
- Cells were harvested, total RNA extracted and the 5' RLM RACE assays conducted.
- 5' RACE amplicons were verified by agarose gel and Bioanalyzer analysis.
- Two different PCR amplification Master Mixes were evaluated and two different NGS library preparation kits for Ion Torrent library preparation were evaluated.
- NGS libraries were sequenced on the Ion Torrent PGM.
- Data filtering options for selecting the specific target sequences from the dataset were explored.
- Finally, the RACE-Seq datasets were analysed using online bioinformatics tools, command line tools and a custom-built Bowtie aligner pipeline.

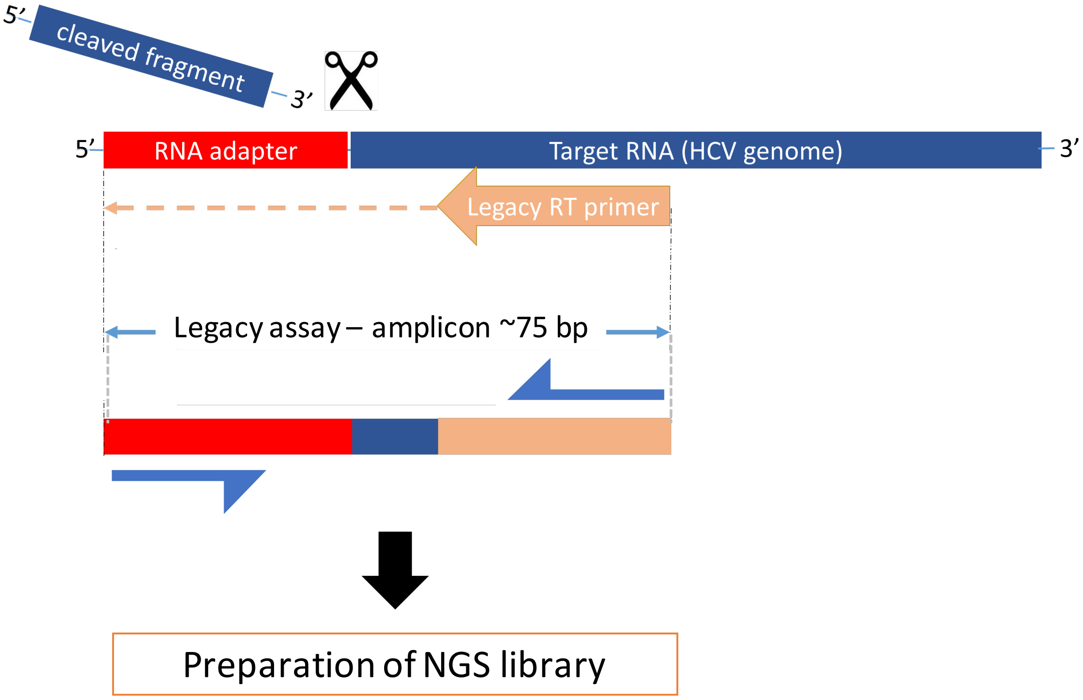


Figure 3.1 Schematic representation of the Legacy RACE-Seq assay.
AGO2-RISC cleavage leaves a phosphate at the new 5' end of the downstream fragment of the cleaved RNA. This end is ligated to a short RNA molecule (the RNA adapter). After reverse transcription with a gene specific primer, anchored PCR is performed with an RNA adapter primer and the reverse transcription primer, generating a short (~75 bp) amplicon.

3.2 Results

3.2.1 Growth and cell morphology of Huh7/con1b cells in culture

Huh7 cells remain the most permissive cell line for maintaining the replication of HCV viral RNA (Blight, et al., 2002; Wakita, et al., 2005; Sainz, et al., 2009; Che, et al., 2012; Wose Kinge, et al., 2014;Yu, et al., 2014). The huh7/con1b cell line used in this study contains the self-amplifying I389/NS3-3'/LucUbiNeo-ET subgenomic replicon, kindly gifted from the lab of Dr Ralf Bartenschlager (Lohmann et al., 1999). The replicon holds a number of adaptive mutations that allow increased accumulation of HCV replicon RNA under G418 selection. The Huh7/con1b cell line has been used extensively for screening of small molecule drugs with the potential to inhibit HCV RNA replication (Devogelaere et al., 2012), to optimise drug antiviral potency (Lu et al., 2004), and to elucidate drug mechanism of action (Te et al., 2007) . The replicon system allows for the stable expression of HCV non-structural proteins.

Huh7 cells were first isolated from a Japanese man with well differentiated hepatocellular carcinoma. Morphologically, the cells are large and flattened, grow as monolayer islands and exhibit a cuboidal epithelial like cell morphology. They retain a low nucleus-to-cytoplasm ratio, are mono- and bi-nucleate and contain multiple nucleoli and cytoplasmic granules (Figure 3.2 A). A study of eight Huh7 cell lines from different laboratories identified morphological differences between cell lines (Sainz et al., 2009). Most cell lines had typical Huh7 morphology while some showed slight differences such as the formation of tightly packed multi-layered islands and differences in cell size. Infecting the cell lines with cell-culture derived HCV (HCVcc) induced HCV-mediated cytopathic effect (CPE) including altered cell morphology, where some cells showed high nucleus-to-cytoplasm ratio resulting in an amoeboid-like shape (Sainz et al., 2009). In the replicon cell line used in this study, this type of CPE was readily observed during cell passage (Figure 3.2 B). Over time, these amoeboid-like shaped cells and other CPE such as the formation of syncytia (polykaryotes) became more announced with successive cell passage. The observation of increased accumulation of such CPE was used to identify the cell passage range for knockdown assays. For this work, cells were used up to passage 38 for assays.

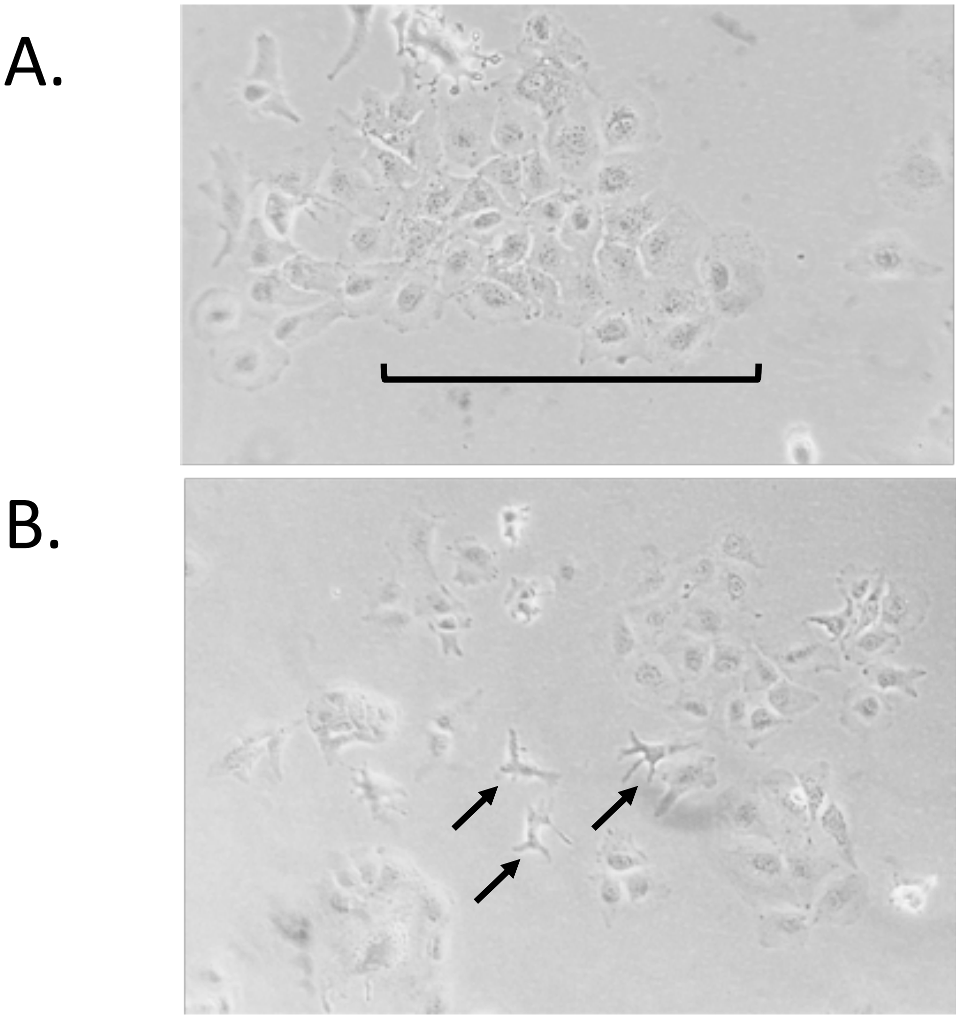


Figure 3.2 Morphology of Huh7/con1b cells maintaining an HCV replicon under G418 selection.

The cuboidal epithelial-like cells grew as attached monolayer islands (A). HCV induced cytopathic effects were typically observed as indicated showing amoeboid-like shaped cells with a high nucleus to cytoplasm ratio (B). (magnification, 100X)

3.2.2 Validation of lower detection limit for luciferase reporter cell line

The high sensitivity of the Luciferase Reporter Gene Assay is a simple, robust and scalable means of assessing specific targeting and suppression of a gene when small double-stranded RNAs are introduced. The linear range of light detection for the Glowmax Multi Detection System (Promega, Madison, USA) was determined by producing a standard curve of luminescence verses relative enzyme concentration. Huh7/con1b cell numbers ranging from 30 cells to 30,000 cells was assessed on a plate-based assay using BRITELite Plus reagent. (See methods Section 2). Two independent attempts indicate a lower limit of quantification of ~300 cells per well with a lower limit of detection of ~30 cells per well (Figure 3.3), identifying that experiments with >10,000 cells per well was a suitable baseline to reliably discriminate up to 99% knockdown.

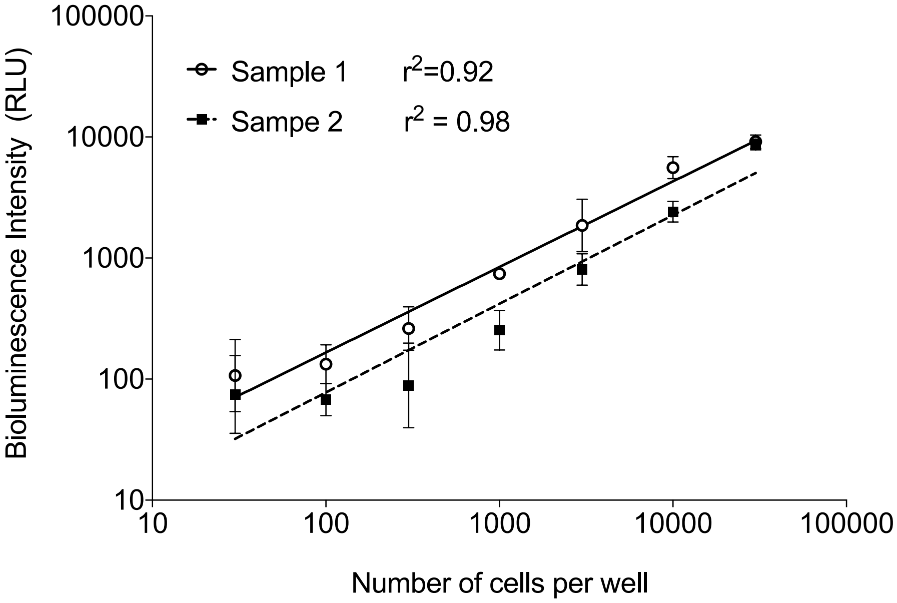


Figure 3.3 Luminescent signal (RLUs) vs fold dilution across a dynamic range spanning 5-folds of magnitude.

A linear relationship is maintained between luciferase signal and cell number even down to less than 100 cells. (Error bars represent n=3)

3.2.3 Potent RNAi by siRNA22 and shRNA22

Knockdown activity for molecules triggering RNAi in a dose dependent analysis of siRNA22 and shRNA22, the highest three concentrations trialled for both siRNA22 and shRNA22 produced over 90% knockdown (Figure 3.4 A). siRNA22 and shRNA22 are designed to target the same region on the HCV replicon genome. As previously noted for shRNA (McAnuff et al., 2007; Vlassov et al., 2007), shRNA22 showed higher potency compared to siRNA22, with EC_{50} for shRNA22 at 5.38 pM and EC_{50} for siRNA22 at 18.9 pM, for this set of dose response assays. Additionally, for all concentrations of siRNA22 and shRNA22, cell viability remained above 80 % when measured by MTT assay (Figure 3.4 B).

From these results, 0.5 nM effective concentration was used for both siRNA22 and shRNA22 for the 5' RACE assays. When measured by the luciferase assay, siRNA22 achieved 96.3% knockdown (Figure 3.5 A) and shRNA22 achieved 94.8% knockdown (Figure 3.5 C) compared to the non-specific oligonucleotide (NSO) control. Cell viability was not affected by transfection of RNAi modalities or other transfection reagents (Figure 3.5 B and D).

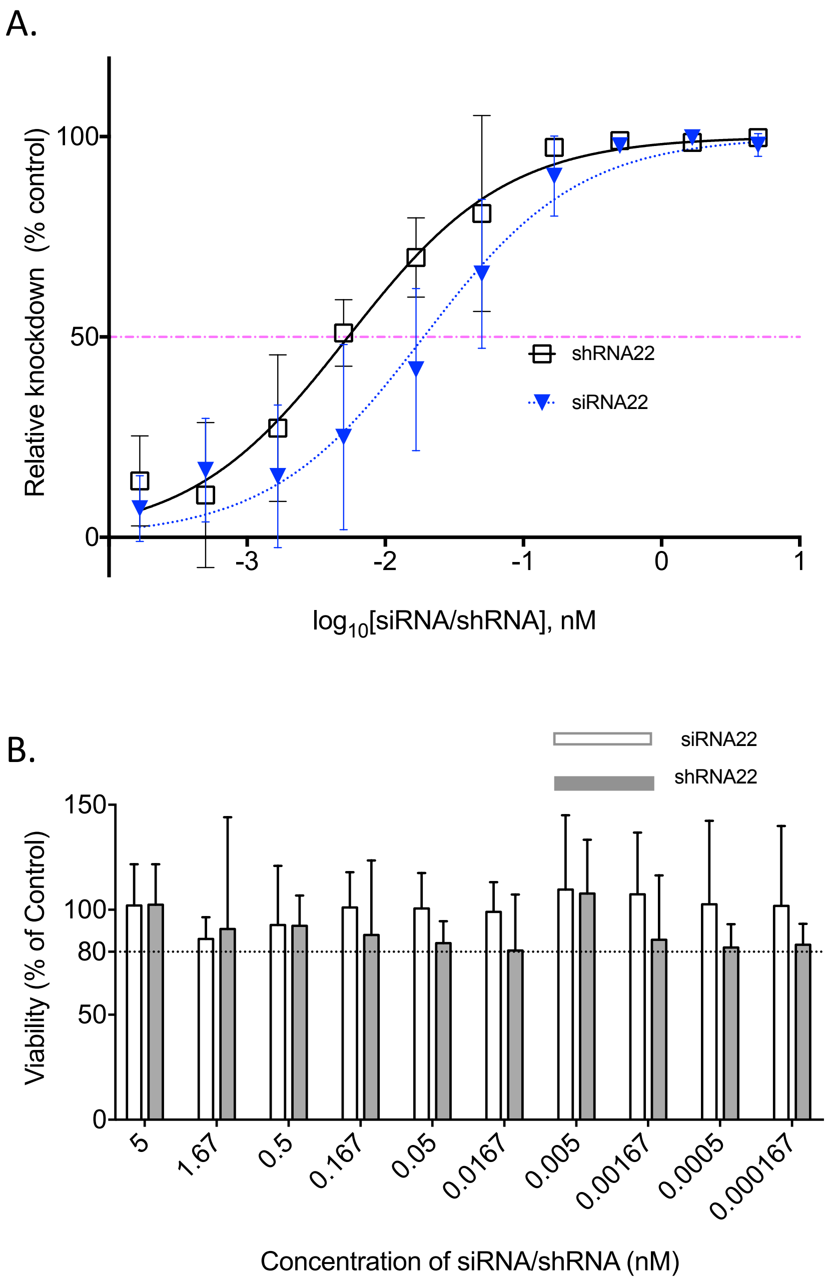


Figure 3.4 Potency of siRNA22 and shRNA22 in the Huh7/Con1b cell line. Huh7/con1b cells were treated with either siRNA or shRNA ranging from 5nM to 167fM. (A) Dose response curves for siRNA22 and shRNA22. Calculation of the half maximal effective concentrations (EC_{50}) found that shRNA22 was more potent than siRNA22. (B) Cell viability was determined by the MTT assay with results indicating <25% toxicity for any of the oligonucleotide concentrations. The mean and standard deviation of n=6 measurements are shown of two independent experiments siRNA22 EC_{50}= 18.95 pM, shRNA22 EC_{50}= 5.38 pM (p=<0.0001), Non-linear regression.

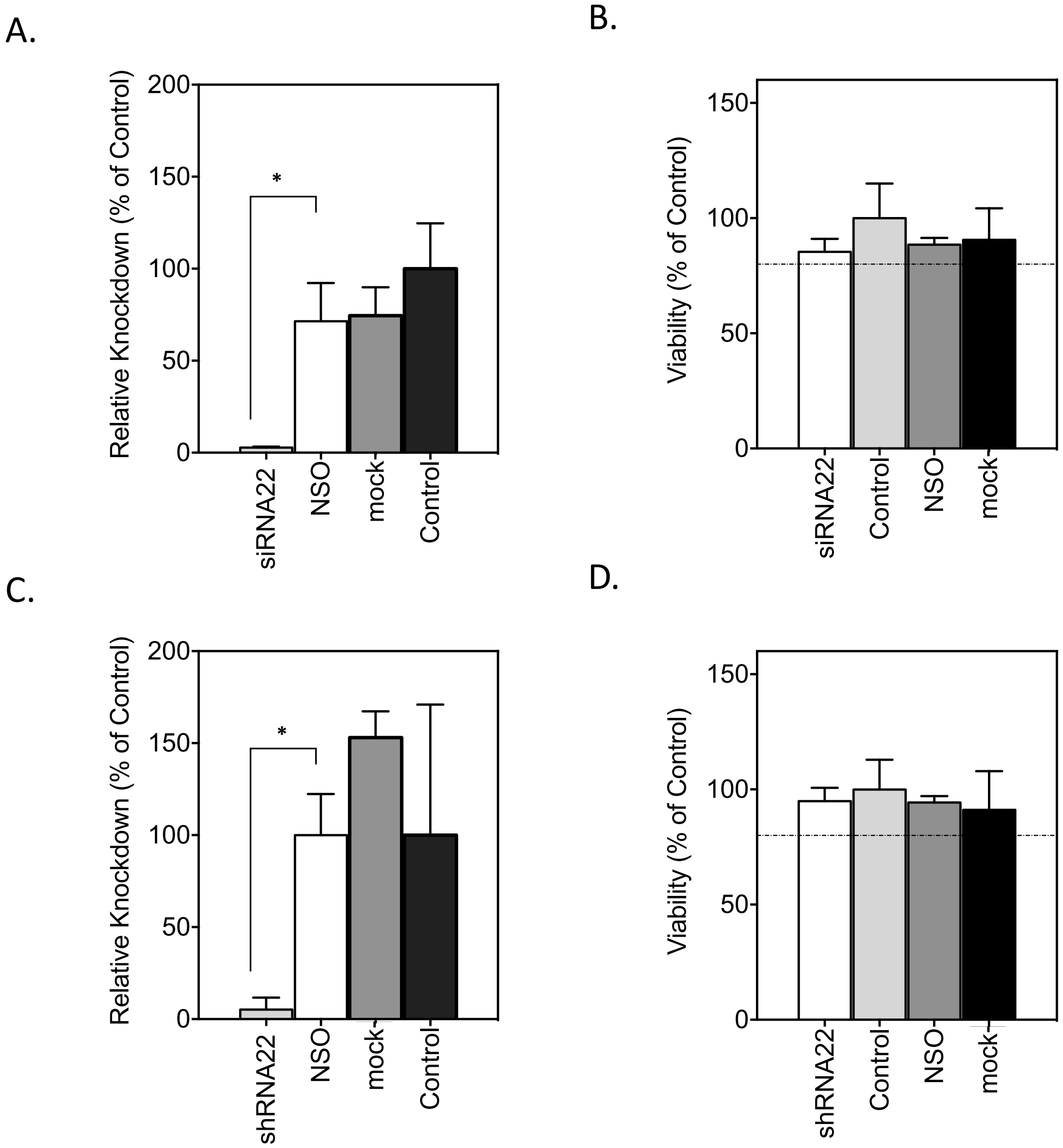


Figure 3.5 Suppressive activity of RNAi analogues.

Cells were transfected with 0.5 nM shRNA22 or siRNA22 and luciferase activity and viability determined after 48 hours. Luciferase knockdown was measured as 96.3%±0.5 for siRNA22 (A) and 94.8%±6.5 for shRNA22 (B) compared to the NSO (non-specific oligonucleotide) control (**P*<0.05 by 2-tailes Paired *t* test). The mean and standard deviation of n=3 measurements is shown. Cell viability was determined by the MTT assay and did not exceed >25% toxicity for siRNA22 (B) or shRNA22 (D). control = no oligonucleotide and no transfection reagent. Mock = no oligonucleotide but transfection reagent included. NSO = non-specofic oligonucleotide and transfection reagent.

3.2.4 RNA extraction, 5’ RLM RACE assays and NGS library preparation

3.2.4.1 Extraction of RNA from knockdown assays

For RNA extraction from 96 well plates, cells from 24 wells on treated plates were pooled by detaching the cells with TrypLE Express and then removing cells from the plate and collecting the cell pellet in a 15 ml centrifuge tube. Total RNA was isolated using the PureLink RNA Mini kit, yielding 71.4 ng/µl RNA for siRNA22 and 74.4 ng/µl RNA for shRNA22 (Table 3.1). Purity of the RNA extractions was measured as A_{260}/A_{280} 2.04 and 2.07 for the respective RNA extractions. RNA was quantified using the Qubit High Sensitivity RNA Reagent kit prior to preparation of 5’ RLM RACE assays.

3.2.4.2 5’ RLM RACE assays and NGS library preparation: Impact of choice of polymerase and library preparation kit

5’ RACE assays were prepared by ligating ~450 ng of RNA to the GeneRacer RNA adapter. After purification of the adapter ligated RNA, the gene specific primer for HCV replicon sequence, primer 22+2 was used to generate the template cDNA. Subsequent amplification used the Platinum Hot Start PCR Master Mix and primers GeneRacerF1 and primer 22+2 (Table 2.3). The 5’ RACE-PCR amplicons were verified on a 3% agarose gel (Figure 3.6). The expected amplicon size was detected, but substantial smear was observed around the expected position when visualised on agarose gel, indicating possible non-specific amplification. Since agarose gel analysis could not indicate a defined peak profile, amplicons were also assessed by Bioanalyzer High Sensitivity DNA chip, which identified that indeed, the expected amplicon fragment size range, which was approximately 70-85 bp compared to the full range of fragments observed which ranged from 50-200 bp occupied only 16% and 17% of fragments in the PCR sample for siRNA22 and shRNA22 respectively (Figure 3.7 A and B). Preparation of NGS libraries using the Ion Plus Fragment Library Preparation kit (Thermo Fisher Scientific) resulted in the desired fragments being 18% and 14% of all fragments in the NGS samples for siRNA22 and shRNA22 respectively (Figure 3.7 C and D). Subsequently, after size selection using the LabChip XT instrument, the amount of desired fragment size was reduced to only 10% and 6% for respective samples (Figure 3.7 E and F). Analysis of the fragment profiles produced by Bioanalyzer analysis and past experience sequencing amplicon libraries on the Ion Torrent PGM strongly indicated that the libraries were of poor quality and were likely to yield insufficient data.

The Ion Plus Fragment Library Preparation kit requires a total of four bead clean rounds when preparing the RACE-Seq samples, which probably contributed to the loss of desired fragment range. The NEBNext Fast DNA Library Prep Set^{TM} protocol on the other hand does not have a bead clean step between the end repair step and the adapter ligation step (Figure 3.8), which would help to retain the desired fragment range. Additionally, the NEBNext kit is much cheaper, reducing the cost per sample for library from ~£50 per sample using the Ion Torrent kit to ~£13 using the NEBNext kit. The NEBNext kit uses Q5 High Fidelity DNA polymerase as part of the protocol, so this polymerase was trialled for amplification of the 5' RACE cDNA. Amplification with the Q5 Hot Start High-Fidelity 2X Master Mix generated a peak profile where the desired fragment range for siRNA22 was 51% and shRNA22 was 57% (Figure 3.9A and B). After library preparation, the desired fragment range (145-165 bp) was assessed to be around 33% for each of the samples (Figure 3.9 C and D) but after LabChip size selection the desired fragment range for sample siRNA22 was 44% of the completed library and sample shRNA22 had at least 50% of the completed library within the desired fragment range (Figure 3.9 E and F).

For library preparation, the PCR reactions were bead cleaned and quantified using the Qubit High Sensitivity DNA kit, with samples generating 4.68 ng/µl DNA for siRNA22 and 5.92 ng/µl for sample shRNA22 (Table 3.1). Subsequently, library preparation used 25.5 µl volume of each cleaned PCR reaction for library preparation resulting in 119 ng and 150 ng as DNA input (Table 3.1). After end repair and adapter ligation reactions, the desired fragment size increased from ~74 bp by a further 84 bp due to addition of the sequencing adapters. The adapter ligated samples were bead cleaned and then amplified in 7 cycles of PCR to enrich for correctly orientated sequencing adapters. The final library was quantified as 56.8 ng/µl for siRNA22 and 38.0 ng/µl for shRNA22 (Table 3.1).

Table 3.1 Nucleic acid quantification at different stages of RACE-Seq sample preparation.

Sample	RNA Input to 5’RLM RACE	[PCR reaction] (ng/µl)	DNA input for NGS library preparation	[NGS library] (ng/µl)
siRNA22	428.4 ng	4.68	119.3 ng	56.8
shRNA22	446.4 ng	5.92	151 ng	38.0

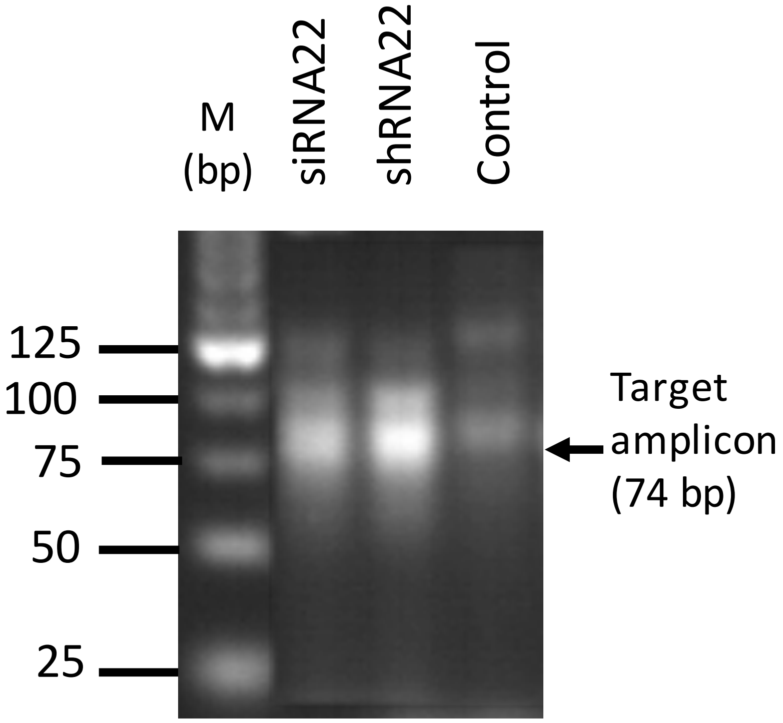


Figure 3.6 5’ RACE-PCR analysis on 3% agarose gel.

Expected amplicon size is ~74 bp. No specific amplification was detected for an untreated control sample. (M=molecular ladder, 25 bp ladder)

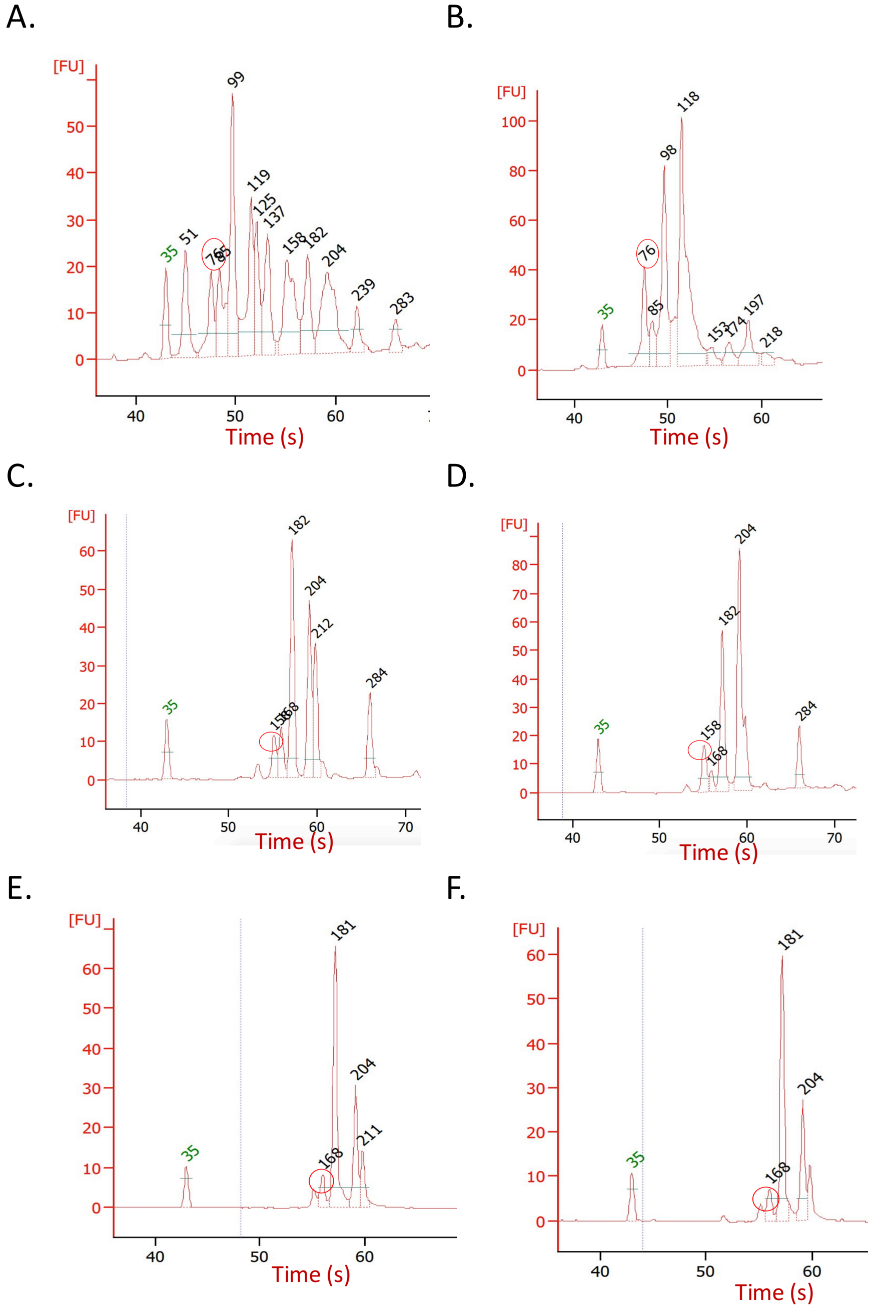
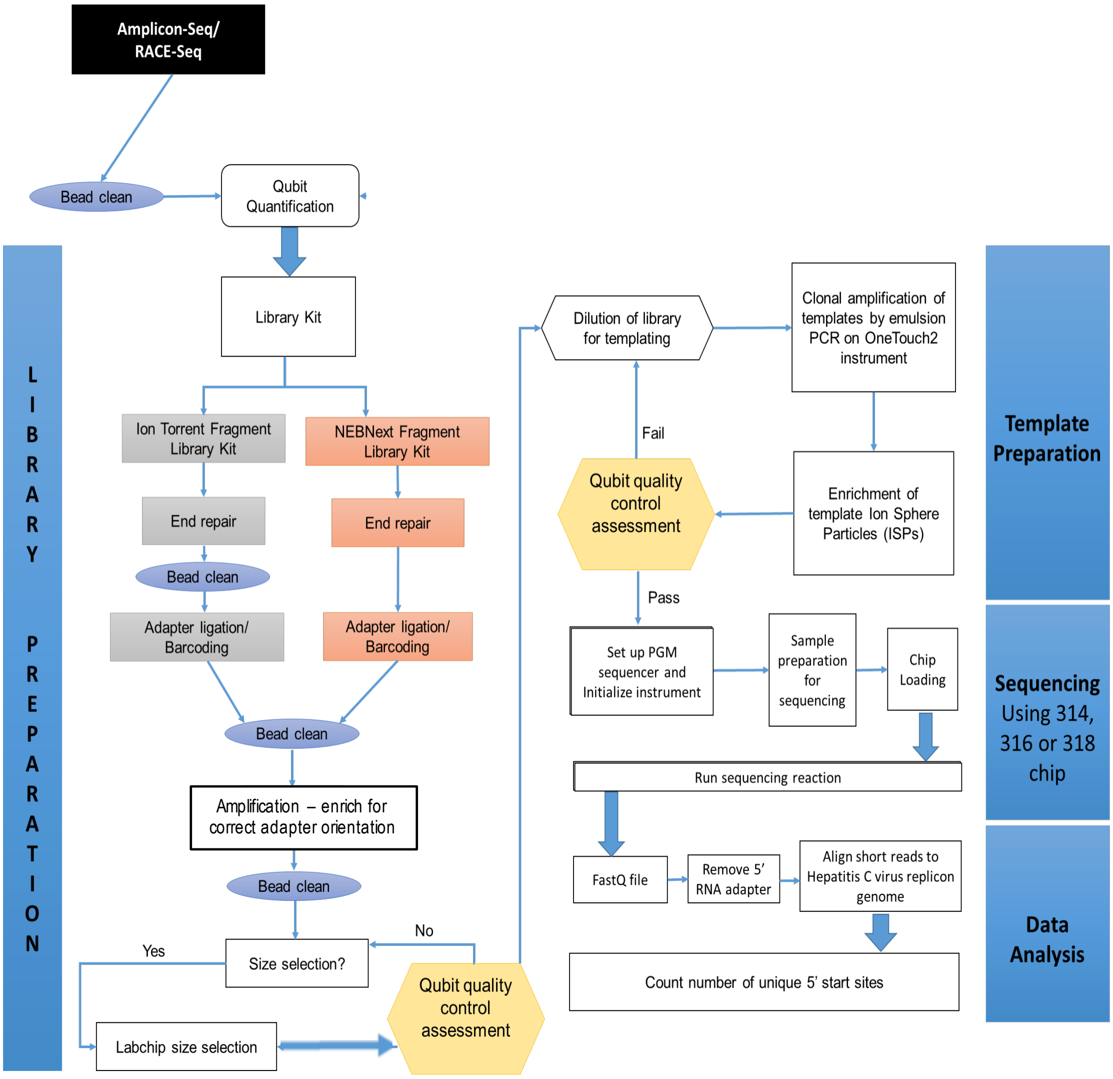


Figure 3.7 Failed library preparation using Platinum Taq Polymerase and Ion Plus Fragment Library Preparation kit.

5’ RACE cDNA was amplified using Platinum Taq Hot Start Master Mix generating a desired fragment range (70-85 bp) of 16% and 17% for siRNA22 (A) and shRNA22 (B) respectively for total PCR fragments. After library preparation, the desired fragment range (145-165 bp) was only 18% and 14% for siRNA22 (C) and shRNA22 (D) respectively. Size selection using the LabChip XT instrument reduced the desired library size range to 10% and 6% for siRNA22 (E) and shRNA22 (F). Bioanalyzer graphs are plots of Fluorescence intensity [FU=fluorescent units] which is influenced by the amount of sample in each chip well vs migration time in seconds (s) for each sample. The desired fragment size is circled in each case.



Amplicon-Seq/ RACE-Seq
Bead clean
Qubit Quantification
L I B R A R Y P R E P A R A T I O N
Library Kit
Ion Torrent Fragment Library Kit
NEBNext Fragment Library Kit
End repair
End repair
Bead clean
Adapter ligation/ Barcoding
Adapter ligation/ Barcoding
Bead clean
Amplification – enrich for correct adapter orientation
Bead clean
Yes
No
Size selection?
Labchip size selection
Qubit quality control assessment
Dilution of library for templating
Clonal amplification of templates by emulsion PCR on OneTouch2 instrument
Fail
Qubit quality control assessment
Enrichment of template Ion Sphere Particles (ISPs)
Pass
Set up PGM sequencer and Initialize instrument
Sample preparation for sequencing
Chip Loading
Run sequencing reaction
FastQ file
Remove 5' RNA adapter
Align short reads to Hepatitis C virus replicon genome
Count number of unique 5' start sites
Template Preparation
Sequencing Using 314, 316 or 318 chip
Data Analysis

Figure 3.8 Full outline of RACE-Seq library preparation, sequencing and data analysis workflow.

The workflow for the Ion Torrent Fragment Library kit (grey) and the NEBNext Fragment Library Kit (orange) is depicted on the left. After size selection, libraries are diluted and pooled for emulsion PCR. The templated library is then sequenced on the Ion Torrent PGM. The FASTQ data file output is pre-processed and aligned to the reference sequence and the 5' end counts generated.

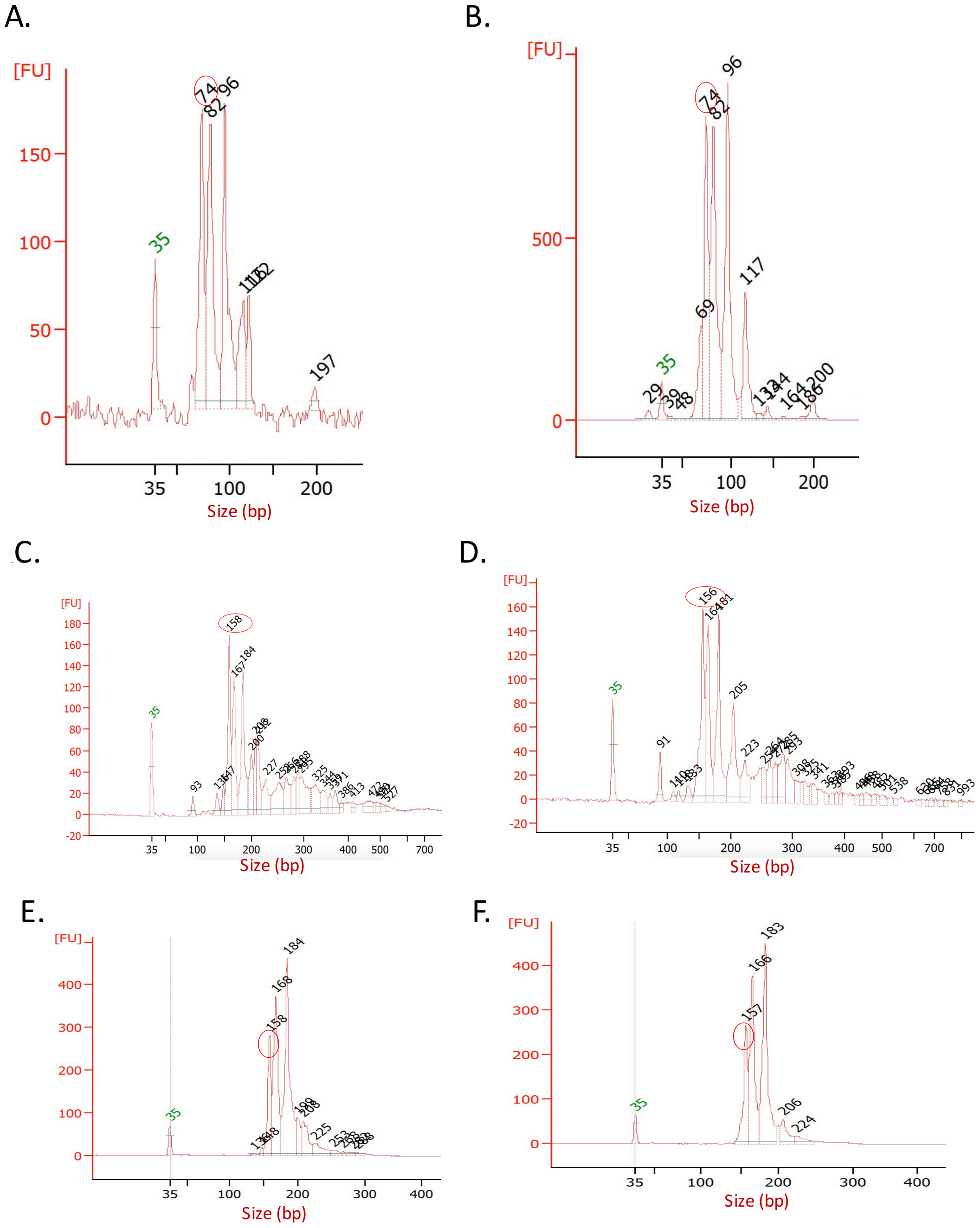


Figure 3.9 Successful RACE-Seq library preparation using Q5 DNA polymerase PCR mix and the NEBNext Library Prep Set for Ion Torrent.

5' RACE cDNA was amplified with Q5 Hot Start 2X Master Mix to generate amplicons where the desired fragment range (70-85 bp) consisted of at least 51% and 57% for siRNA22(A) and shRNA22(B) respectively. After library preparation, the desired fragment range (145-165 bp) was about 33% for siRNA22(C) and shRNA22(D). Size selection using the LabChip XT instrument increased the desired library size range to 44% for siRNA22(E) and 50% for shRNA22 (F). (Bioanalyzer graphs are plots of Fluorescence intensity [FU=fluorescent units] which is influenced by the amount of sample in each chip well vs fragment size (bp=base pairs). The desired fragment size is circled in each case.

3.2.5 Size selection as a critical component of library preparation

The Ion Torrent System uses an automated emulsion PCR instrument (Ion One Touch2 instrument) to clonally amplify single templates to special sequencing beads. The reagents and protocol are proprietary with all validated protocols from Ion Torrent based on discrete fragment insert size of either 100 bp, 200 bp or 400 bp. The standard library preparation is for 200 bp inserts with templates larger than this range prone to poor templating. Thus, size selection of 5’ RACE amplicons was critical to enable a discrete fragment size range for input into emulsion PCR. Size selection also enhanced the fraction of desired fragment size. The LabChip XT instrument is an automated fragment analyser capable to partitioning a discrete fragment range for collection. The LabChip XT was set at 158 ± 10% to collect the fraction of fragments around the expected library size (Figure 3.10 A). Size selection enriched for the desired fragment size and eliminated larger unwanted fragment sizes from the library preparations (Figure 3.10 B). The size selected libraries were quantified on Bioanalyzer DNA chip and diluted for input to emulsion PCR. After emulsion PCR, samples were cleaned using the Enrichment system and each sample sequenced on an individual 318 V2 chip on the Ion Torrent PGM.

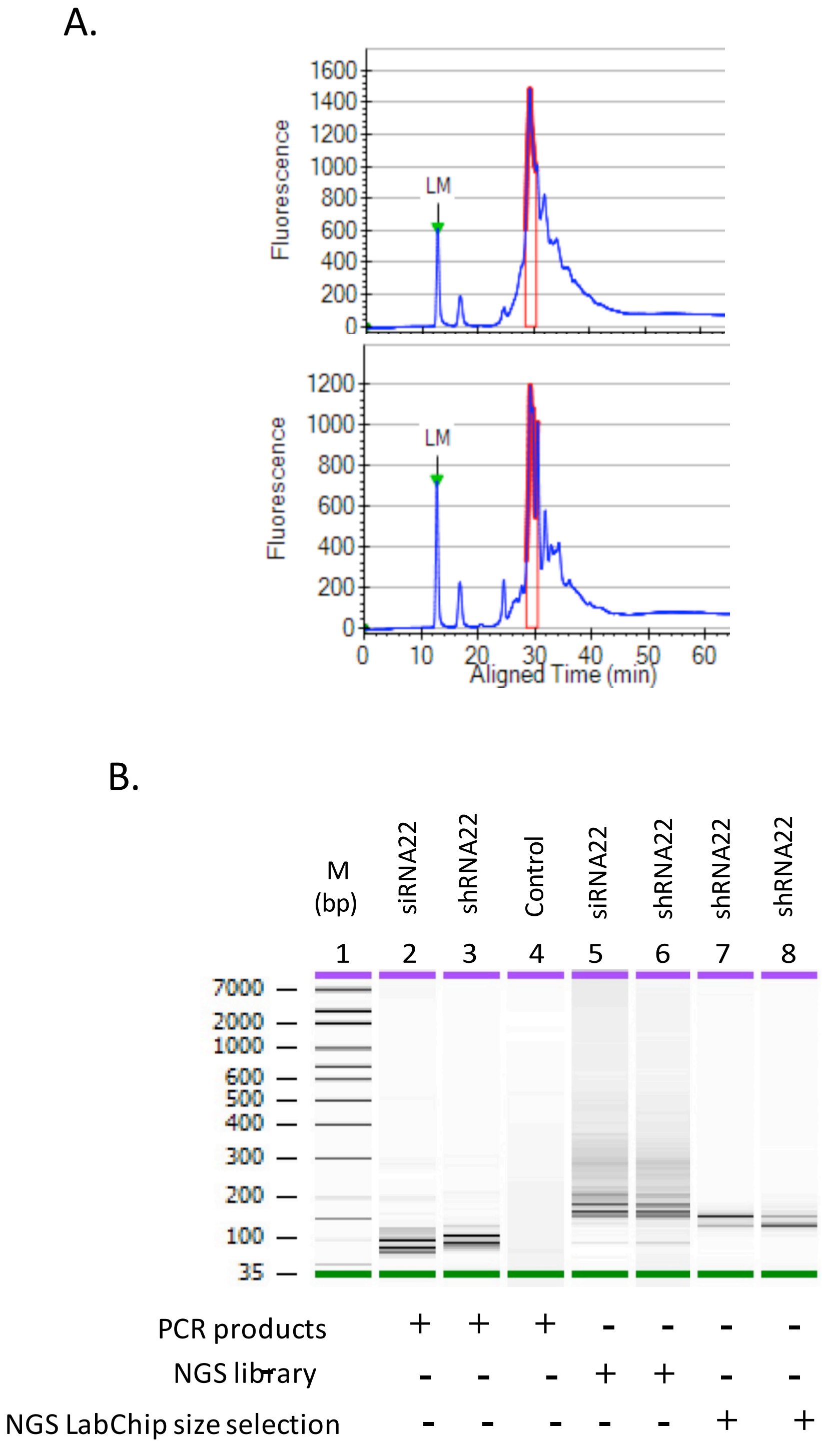


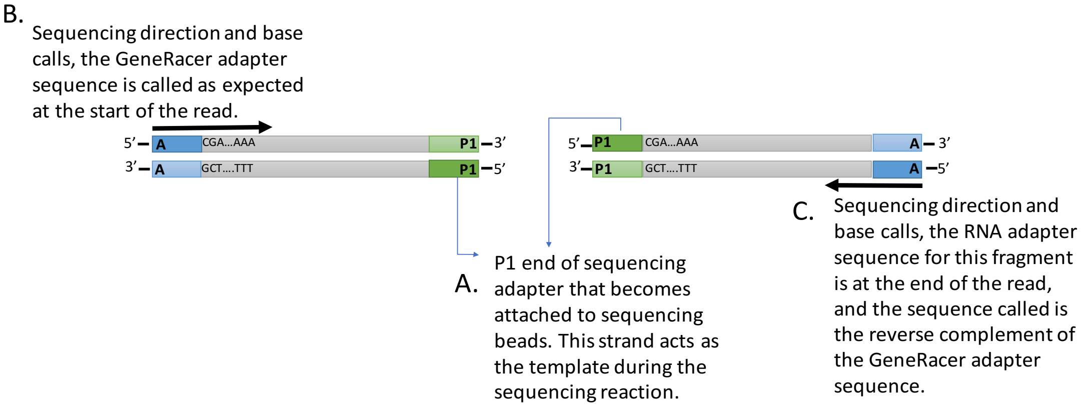
Figure 3.10 Impact of LabChip size selection on isolation of desired fragment size for RACE-Seq library preparation.

The LabChip automated size selection system partitions a discrete size fraction (in red outline) for collection (A). After 5’ RACE PCR (B lanes 2, 3 and 4), NGS library preparation captures and amplifies some non-specific products (B lanes 5 and 6), which were eliminated by size selection using the LabChip XT instrument (B lanes 7 and 8). (M= molecular ladder)

3.2.6 RACE-Seq data analysis

3.2.6.1 The GeneRacer adapter sequence is identified in both orientations

During library preparation, the A and P1 sequencing adapters ligate to amplicons in four different configurations (Figure 1.7). The subsequent amplification step post adapter ligation enriches for A-amplicon-P1 or P1-amplicon-A ligations which can occur equally in either orientation. The P1 end of the library fragments become templated to the sequencing beads. Later, the sequencing reaction occurs in the 5’ to 3’ direction, resulting in the GeneRacer adapter sequence occurring either at the beginning of the sequenced read or at the end of the read (Figure 3.11). Understanding the orientation of the GeneRacer adapter sequence in the sequenced dataset was important in designing the first step in analysing the data. By searching for the GeneRacer adapter sequence at both ends of the reads ensures comprehensive filtering of the dataset and prevents unnecessary loss of usable data.


B. Sequencing direction and base calls, the GeneRacer adapter sequence is called as expected at the start of the read.

A. P1 end of sequencing adapter that becomes attached to sequencing beads. This strand acts as the template during the sequencing reaction.

C. Sequencing direction and base calls, the RNA adapter sequence for this fragment is at the end of the read, and the sequence called is the reverse complement of the GeneRacer adapter sequence.

Figure 3.11 Illustration to indicate how the RNA adapter sequence is output in both the forward and reverse complement directions.
The A and P1 adapters ligate to 5’ RACE amplicons as Adapter-PCR product-Adapter, resulting in a mix of configurations where the A or P-adapter can ligate upstream or downstream of the GeneRacer sequence (CGA…AAA) within the PCR amplicon. In emulsion PCR, the P1-adapter end becomes templated to the sequencing beads (A). This strand acts as the template in the sequencing reaction. Sequencing occurs in the 5’ to 3’ direction, resulting in the adapter sequence occurring either in the forward direction (B) or the adapter sequence occurs at the end of the sequenced read as the reverse complement of the RNA adapter sequence (C).

3.2.6.2 RACE-Seq data output from the Ion Torrent PGM

On the completion of the sequencing run, a run report is generated. The run report provides information regarding all aspects of the data analysis. Besides the statistics on the number of templated beads detected on the chip (loading) and loss of data due to standard filtering operations (Figure 3.12), samples are de-multiplexed and total number of reads per sample reported.

3.2.6.3 Filtering of RACE-Seq datasets and alignment to the reference

Read output from the Ion Torrent Sequencer was approximately 3 million reads for each sample. For siRNA22, 47.1% of reads had the GeneRacer adapter sequence in the reverse orientation while this was 43.4% for the shRNA22 sample (Table 3.2). Reverse complementing these datasets and concatenating the reads with the set of data with the expected GeneRacer adapter sequence in the forward direction, captured 90.2% and 92.0% of all usable reads for siRNA22 and shRNA22 respectively (Table 3.2). Further data processing steps included removal of the GeneRacer adapter sequence, filtering for reads that held the amplification primer sequence on the 3' end, trimming of low quality bases from the 3' end of reads and removal of short reads (<30 bp). This final data set constituted 65.5% (1,932,233 reads) and 76.0% (2,377,583 reads) of all usable reads for siRNA22 and shRNA22 respectively. This filtered dataset was aligned to the HCV replicon genome using an adapted version of the RACE-SEQ-LITE pipeline (Theotokis et al., 2017). The 5' end counts that were generated were then (1) plotted at linear scale as number of 5'ends at positions within the expected RISC hybridization site (Figure 3.13 A and C) and (2) the 5' end counts were calculated as a percent of total aligned reads and plotted as the percent of aligned reads within the expected RISC hybridization site (Figure 3.13 B and D).

For siRNA22, 22.1% of filtered reads aligned to the HCV replicon reference when zero mismatch tolerance was implemented. Further analysis showed that 86.9% of all aligned reads for siRNA22 were within the expected RISC hybridization site (Table 3.2), with the remaining reads spread out at regions outside of this area. For shRNA22, 25.1% of all filtered reads aligned to the HCV replicon reference sequence, with 94.3% of aligned reads at the expected RISC hybridization site (Table 3.2). An unexpectedly large number of filtered reads for each data set failed to align to the HCV replicon reference, with 77.9% and 74.9% of reads partitioned to the ‘unaligned data’ output file for the siRNA22 and shRNA22 samples respectively.

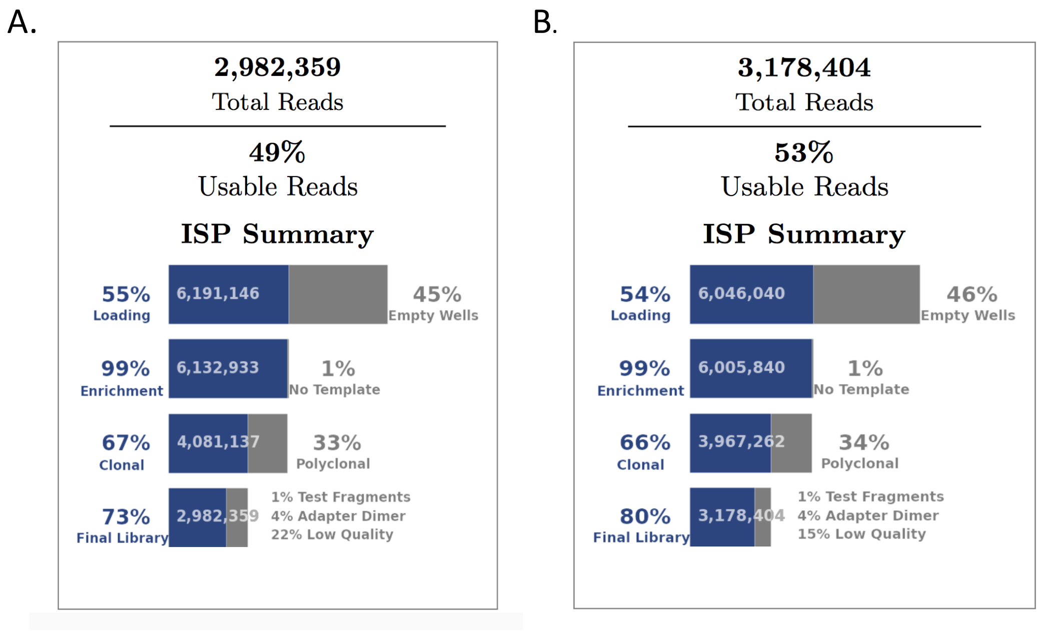


Figure 3.12 Ion Torrent PGM summary of data for legacy RACE-Seq assays. Each RACE-Seq sample generated approximately 3 million reads. (A) siRNA22 sample has 55% chip loading which equated to 6,191,146 templated beads and (B) shRNA22 has 54% chip loading which equated to 6,046,040 templated beads. The Ion Torrent server filtered each of the datasets for polyclonal reads, low quality reads, and adapter dimer to give the final output of usable reads.

Table 3.2 RACE-Seq data analysis for siRNA22 and shRNA22 (legacy assays)

	siRNA22		shRNA22	
	Legacy RACE-Seq data		Legacy RACE-Seq data	
Filter Criteria	**Read count**	**Percentage**	**Read count**	**% of Total reads**
Total number of reads	2948412		3129245	
Total reads with GeneRacer adapter	1270582	43.1*	1519754	48.6*
Total reads with inverted GeneRacer sequence	1388602	47.1*	1358774	43.4*
Total reads with GeneRacer adapter	2658892	90.2*	2878141	92.0*
Reads remaining after filtering	1932233		2377583	
Total reads that align to HCV reference	427070	22.1**	596359	25.1**
Total reads that align to target site	371240		562369	
Total unaligned reads	1505163	77.9**	1781224	74.9**

* Percent of total reads
** Percent of filtered reads

3.2.7 Interpretation of RACE-Seq peak profiles

RACE-Seq analysis identified the expected cleavage site for both the siRNA22 and the shRNA22 sample by reporting the highest number of reads at the expected cleavage site (Figure 3.13 A and C, lightly shaded bar, T=9489). Additionally, for both samples, a substantial number of reads (35.7% and 45.3%) were found downstream of the expected cleavage site (Figure 3.13 B and D) and are presumed to be degradation products arising from 5' to 3' RNA degradation of the RISC cleaved targets and cannot be validated as possible RISC derived cleavages. Upstream of the expected cleavage site, RACE-Seq reported additional 5' ends for both siRNA22 and shRNA22. The siRNA22 sample had 4 peaks above 1% at positions upstream of the expected cleavage site. Of these, position A=9486 (at +3 from expected cleavage site) reported the second highest secondary peak count at 17.0% of aligned reads and position C=9488 (at +1 from cleavage site) reported the third highest secondary peak count at 7.4% (Figure 3.13 B). For the shRNA22 RACE-Seq sample, one additional 5' end peak stood out at C=9482 with 3.6% of aligned reads (Figure 3.13 D). However, this peak position was not corroborated by the siRNA22 RACE-Seq sample.

Sample siRNA22 had 86.9% of all aligned reads with a 5' end within the RISC hybridization site (Figure 3.13 B), while shRNA22 had 96.5% of aligned reads within the RISC hybridization site (Figure 3.13 D). In each case, the expected AGO2-RISC-induced cleavage position was reported as the primary peak. For the additional peaks at positions upstream of the expected cleavage site, whether they are induced by RISC activity remains unclear.

3.2.8 Analysis of unaligned data

Unexpectedly, a large portion of the filtered datasets for siRNA22 and shRNA22 failed to align to the HCV replicon reference sequence (Table 3.3). Even when allowing up to 3 mismatches for the Bowtie aligner, 70.3% of reads for siRNA22 and 68.1% of reads for shRNA22 still failed to align, leading to the conclusion that these reads were either not HCV sequences or, were HCV reads with >3 mismatches to the reference. When the collapsed datasets were mapped to the HCV replicon reference (without reverse primer sequence) a mix of HCV and non-HCV sequences were identified (Figure 3.14). The RML-RACE assay is not specific for AGO2-RISC cleaved products thus the RNA adapter may attach to other transcripts and even the reverse primer may generate spurious cDNA templates.

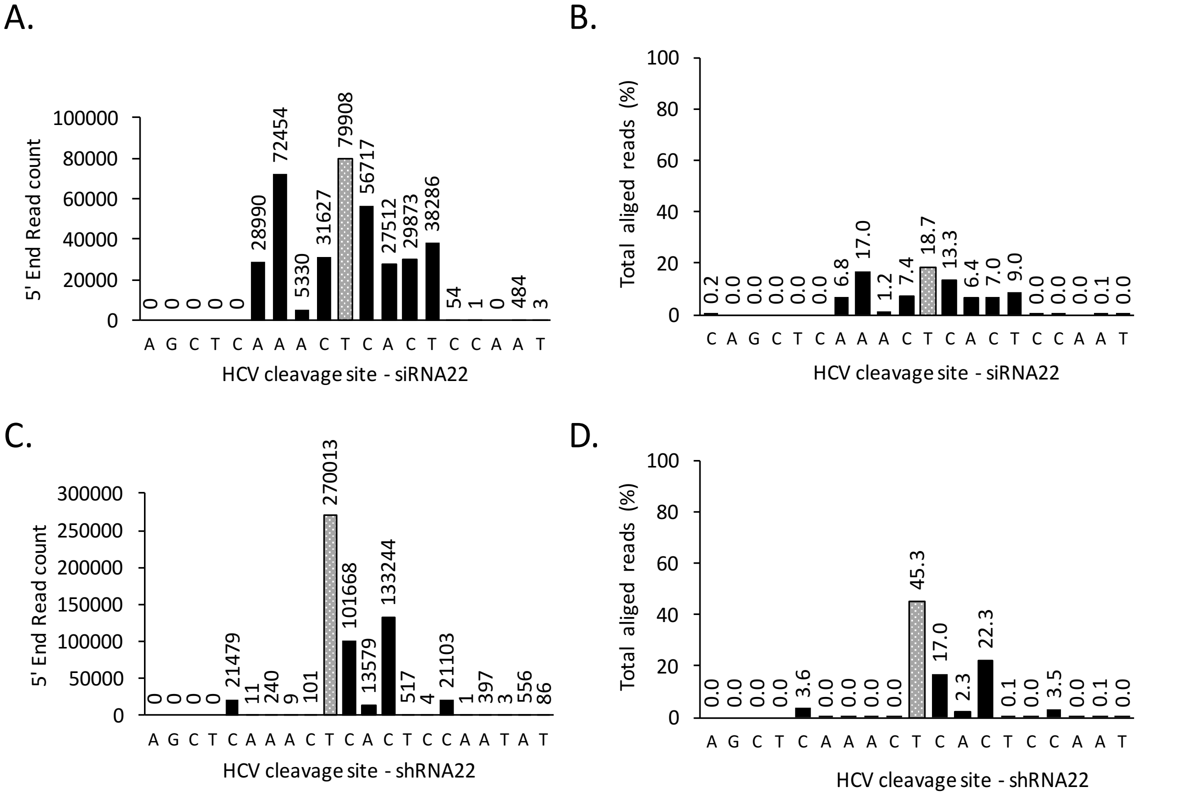


Figure 3.13 Novel 5' ends identified by RACE-Seq legacy assay on Ion Torrent PGM. For both siRNA22 and shRNA22, RACE-Seq reports captured reads with 5' ends around the expected AGO2-RISC cleavage position. The expected RISC cleavage position is indicated at the lightly shaded bar. Graph (A) represents the absolute read count of aligned reads and their 5' end position for sample siRNA22. Graph (B) is the percentage of total aligned reads for each peak within the RISC hybridization site for siRNA22. Graph (C) represents the total number of aligned reads with 5' ends within the cleavage site for shRNA22 with graph (D) indicating the percentage of aligned reads that can be attributed to each peak within the RISC cleavage site for the shRNA22 sample.

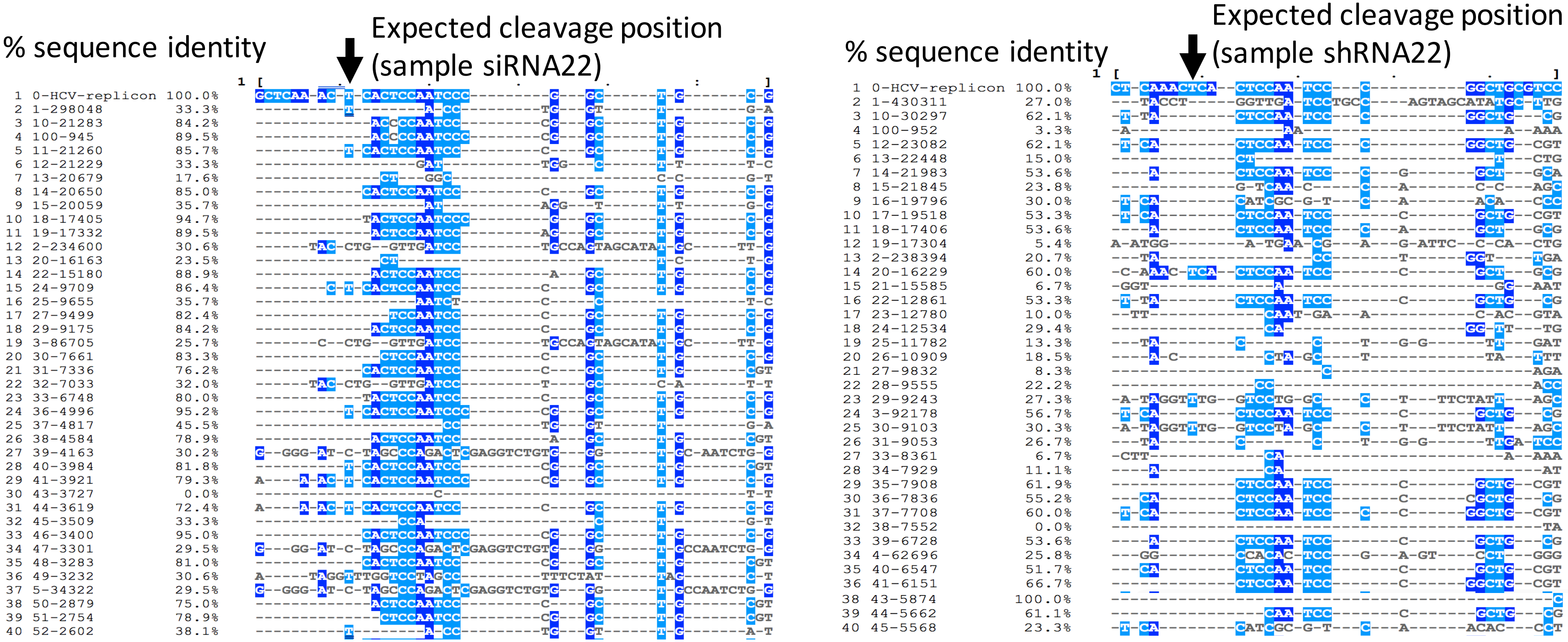


Figure 3.14 Analysis of unaligned reads for siRNA22 RACE-Seq and shRNA22.

Data that did not align to the HCV replicon reference at mm=0 had the reverse primer sequence removed, the reads collapsed and then multiple sequence alignment performed against the HCV replicon reference sequence at target site 22. The top 40 read counts are analysed for sequence similarity to the reference.

3.3 Discussion

This chapter describes the process of transferring a legacy RACE-Seq assay described in Denise et al., (2014) from sequencing on a high throughput platform, the Illumina HiSeq2000 instrument, to sequencing on the low throughput platform, the Ion Torrent PGM. The process revealed a number of critical factors that are generally applicable to RACE-Seq but also some that are particular to sequencing on the Ion Torrent platform. Typically, greater than 2 µg of RNA is reported as input to the RNA adapter ligation reaction for standard 5' RML RACE assays (Zimmermann et al., 2006; Neff et al., 2011; Hagopian et al., 2017). This is largely a carryover from a protocol from one of the most popular 5' RACE Assay Kits, the FirstChoice RML-RACE Kit (Thermo Fisher Scientific). The RACE-Seq assay performed here, was able to reduce the RNA input to ~428 ng for siRNA22 and ~446 ng for shRNA22. This could prove valuable if the assay were to progress to clinical stage use, where a single patient sample i.e., liver biopsy needs to be analysed in multiple assays (Tabernero et al., 2013).

Being able to critically evaluate the fragment size range of the 5' RACE PCR reactions as well as the post library and post size selection samples using the Bioanalyzer chip assay proved to be a critical tool, ensuring the eventual success of the RACE-Seq assays. It allowed performance of the Q5 Hot Start High-Fidelity 2X Master Mix to be evaluated against the poorer performing Platinum Hot Start PCR Master Mix. Additionally, after size selection of the NGS libraries on the LabChip instrument, Bioanalyzer analysis confirmed that there was sufficient amount of the desired fragment size range (44% for siRNA22 and 50% for shRNA22) in the final library preparation to advance the sample to sequencing on the Ion Torrent PGM. The Illumina sequencing platform may not have the same requirement for stringent size selection of libraries that the Ion Torrent PGM does (Section 1.5). When sequencing on the Illumina platform, the length of the sequenced fragment is determined by the number of flows, so that a 400 bp fragment for example can have 150 bp read from one end of the fragment and another 150 bp read from the other end of the 400 bp fragment. Since only the region after the GeneRacer RNA adapter is of interest, not having the central part of the fragment has little consequence.

The LabChip XT instrument allowed a discrete DNA fragment size range to be isolated and very efficiently eliminated larger fragments from the final library preparations. Electronic fragment analysers (such as the Bioanalyzer) and automated DNA size selection instruments (such as the LabChip instrument) are non-standard pieces of laboratory equipment. While their implementation may not be an absolute requirement for RACE-Seq for high throughput NGS systems, certainly they both proved valuable for designing and characterising RACE-Seq assays for the Ion Torrent PGM.

Sequencing the siRNA22 and shRNA22 RACE-Seq samples on the Ion Torrent PGM generated ~3 million usable reads per sample on 318 V2 sequencing chips. In filtering the RACE-Seq datasets for the target sequence, loss of some reads occurred, but the filtering criteria still retained 65.5% of usable reads for siRNA22 and 76% of usable reads for the shRNA22 RACE-Seq sample. Part of this success hinged on an acute understanding of NGS library preparation which highlighted the fact that the RNA adapter sequence could appear at either end of the raw sequence reads. As the first step in filtering the data searches for the adapter sequence, a failure to capture the reverse complement sequence would have resulted in a loss of just under 50% usable reads. For the legacy RACE-seq assay however, after mapping the filtered datasets to the HCV replicon reference sequence and counting the 5' ends, a substantial (77.9% and 74.9% for respective samples) number of reads failed to align to the HCV replicon reference when mismatch of zero was applied. Since the HCV replicon genome actually exists as quasispecies RNA genomes(McWilliam Leitch and McLauchlan, 2013; Park et al., 2014; Chen et al., 2016), the mismatch tolerance was increased to a maximum of 3, but only marginally increased the number of reads that aligned to the reference, with an increase of 7.5% for siRNA22 and 6.8% for shRNA22.

The expected cleavage position for the synthetic siRNA22 analogue and the synthetic shRNA22 analogue was reported for each of the RACE-Seq samples and was as reported by classic 5' RACE and Sanger sequencing by Denise et al., (2014). When the Ion Torrent PGM RACE-Seq profile for siRNA22 was compared to the Illumina deep sequencing RACE-Seq profile reported by Denise et al., (2014), no corroboration of peak incidence was detected beyond reporting the expected cleavage position as the major peak. Further, the read depth achieved in the Denise et al., (2014) RACE-Seq dataset, (>12 million reads reported at the expected cleavage site) resulted in all other 5' peaks that appeared upstream of the expected cleavage site reporting at less than 0.2% of all reads that aligned within the siRNA target site. As with other reports of RACE-Seq datasets (Tabernero et al., 2013; Barve et al., 2015), additional 5' end counts were reported upstream as well as downstream of the expected cleavage site. While the downstream products are designated as putative degradation products, classifying the upstream 5' ends remained more challenging. Since adapter ligation by T4 RNA ligase is not specific for RISC cleaved substrates, all additional peaks captured by the RACE-Seq assay cannot be exclusively attributed to being due to RNAi induced activity. Additionally, since the siRNA22 5'RACE profile from Ion Torrent did not correlate to the Denise et al., (2014) profile, this begins to highlight the influences that assay design can place on obtaining robust, reliable, unbiased data in RACE-Seq assays.

RACE-Seq remains an under-utilized assay, without an established protocol. Primarily, it is being used as a substitute for Sanger sequencing to confirm an expected RNAi behaviour of a designed RNAi-based bioactive (Tabernero et al., 2013; Barve et al., 2015; Ganesh et al., 2016). The Denise et al., (2014) publication attempted to derive new insights into RISC cleavage behaviour using RACE-Seq, concluding with the suggestion that AGO2-RISC may show preference for a second cleavage site at positions 13-15 rather than just position 10-11. However, the RACE-Seq assays for siRNA22 and shRNA22 using the same protocol for generating the 5' RACE amplicons and sequencing on the Ion Torrent failed to support such a hypothesis.

The legacy RACE-Seq assay design generated very short PCR products, mainly due to the limitations of the Illumina sequencing platform at the time. However, short amplification reactions are likely to be highly efficient (Suzuki and Giovannoni, 1996; Arezi, *et al.*, 2003) and result in the reaction reaching saturation conditions, including DNA polymerase inhibition due to saturated templates. The short length PCR amplification as well as the differences in Illumina vs Ion Torrent sequencing platforms is likely to have impacted on the failure to derive any new insights regarding RISC behaviour for these Ion Torrent RACE-Seq datasets.

4 AN IMPROVED METHOD FOR PREPARING RACE-SEQ LIBRARIES

4.1 Introduction

The 5’ RACE assay relies on amplifying an adapter-ligated target using PCR. The expectation is that the final amplified reaction represents the diversity and relative frequency of the original sample. For sequencing on the Ion Torrent PGM, fragment size is a critical component of library preparation. A number of Ion Torrent protocols utilise size selection, either using magnetic beads or electronic size selection instruments such as the LabChip XT instrument. By the end of 2016, the LabChip XT instrument, reagents and support was discontinued. The E-gel 2% size select system was chosen as the replacement as it offered much cheaper per sample cost and allows up to eight samples to be run on a single gel compared to the 3 samples for each LabChip XT chip. However, the E-gel system is not an automated system and as such, size selection is prone to manual handling error.

The chemistries, protocols and design of NGS technologies have greatly improved over the last 5 years, and both Illumina and Ion Torrent technologies accommodate libraries with inserts of at least 200 bp. Primer sets for the long amplicon RACE-Seq assay were designed to have final amplicon size greater than 100 bp. In addition, since the samples were to be pooled, particular care was taken to design 5’ RACE PCR assays with similar sized amplicons. To enrich for the expected template, the reverse primer in PCR amplification was placed at a position internal to reverse transcription primer position (Figure 4.1).

This chapter focuses on two main objectives, first, to design a RACE-Seq assay to be compatible with longer amplicon sequencing and second, to explore multiplexing practicalities for sequencing on the Ion Torrent PGM.

The following approaches were taken in order to achieve the objectives of this chapter:

- Two sets of RACE-Seq assays were conducted at two separate time points.
- The time point (1) sequencing run consisted of samples; siRNA22, shRNA22, siRNA6, shRNA6, siRNA19 and shRNA19. The PCR samples were size selected prior to library preparation.
- The time point (2) run consisted of RACE-Seq assays siRNA22, shRNA22, siRNA6 shRNA6, siRNA19*, shRNA19*, siRNA19 and shRNA19 (Figure 4.2). The siRNA19* and shRNA19* samples were size selected prior to library preparation similar to the time point (1) samples.
- New sets of 5’ RACE primers were designed and tested. Combinations of reverse primer for cDNA synthesis and gene-specific amplification primers were evaluated by assessing 5’ RACE PCR products on 2% agarose gels.
- To allow single round PCR reactions, the amount of RNA input to the RNA adapter ligation reaction was increased for time point (2) samples.
- Two different approaches to obtaining the size selected NGS libraries were carried out. Size selection was done either before library preparation or after library preparation.
- The impact of agarose gel extraction on the final NGS library fragment profile was evaluated by Bioanalyzer analysis.
- Multiple 5’ RACE samples were pooled for NGS sequencing runs and the impact on chip loading and final data output was evaluated.

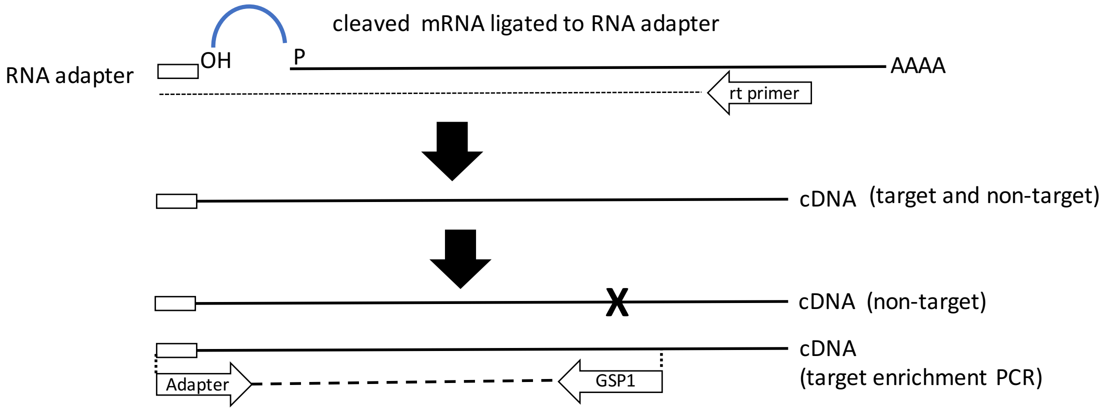


Figure 4.1 Illustration of long amplicon RLM RACE assay.

After adapter ligation and reverse transcription, both target and non-target templates may be generated. Amplification using a gene specific primer internal to the reverse transcription primer specifically enriches for the target of interest.

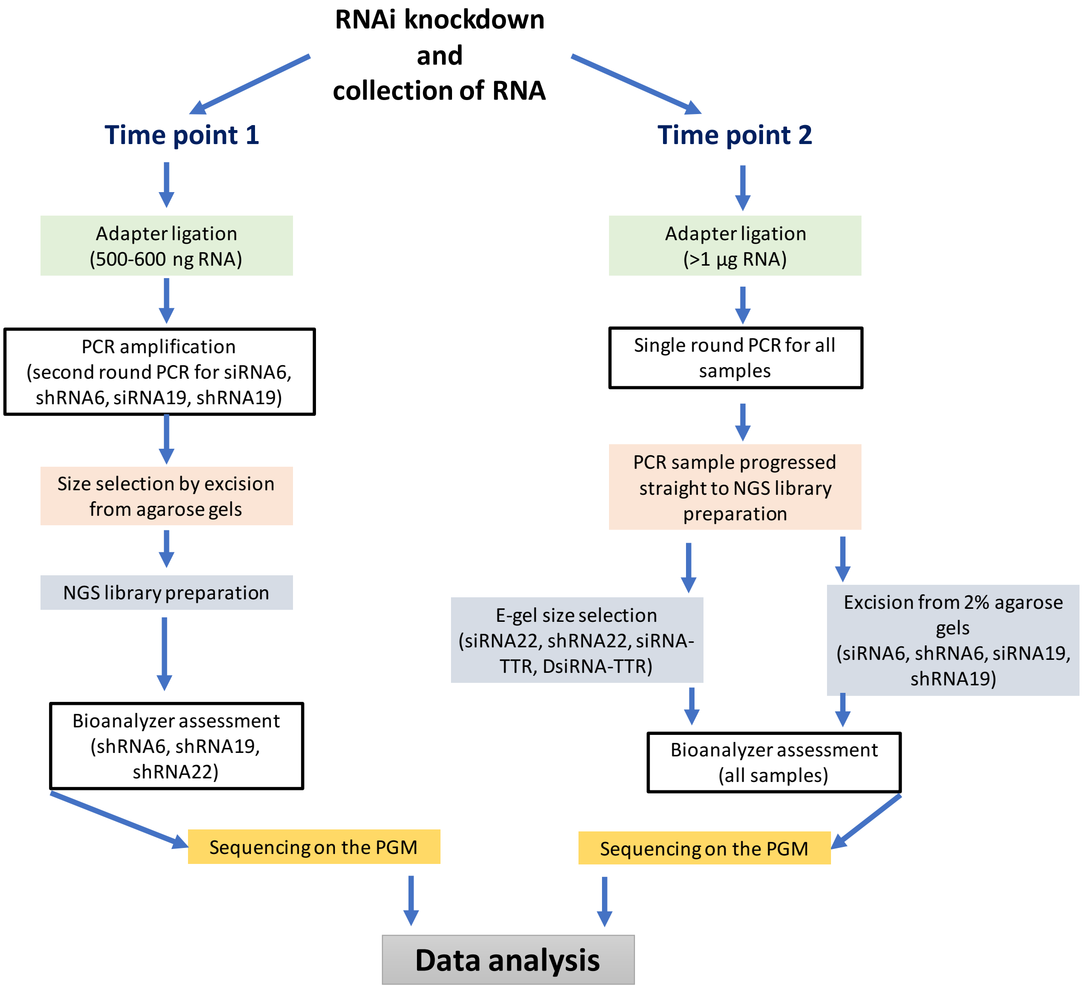

RNAi knockdown and collection of RNA
Time point 1
Time point 2
Adapter ligation (500-600 ng RNA)
Adapter ligation (>1 µg RNA)
PCR amplification (second round PCR for siRNA6, shRNA6, siRNA19, shRNA19)
Single round PCR for all samples
Size selection by excision from agarose gels
PCR sample progressed straight to NGS library preparation
NGS library preparation
E-gel size selection (siRNA22, shRNA22, siRNA-TTR, DsiRNA-TTR)
Excision from 2% agarose gels (siRNA6, shRNA6, siRNA19, shRNA19)
Bioanalyzer assessment (shRNA6, shRNA19, shRNA22)
Bioanalyzer assessment (all samples)
Sequencing on the PGM
Sequencing on the PGM
Data analysis

Figure 4.2 Work flow of RACE-Seq assay for time point 1 and time point 2. This diagram details the sample workflow for two different sets of RACE-Seq assays prepared at separate time points.

4.2 Results

4.2.1 RNAi activity assays

To determine the optimal concentration of siRNA and shRNA to use for the RACE-Seq assays, dose response studies were carried out for siRNA6, shRNA6, siRNA19 and shRNA19. Statistical analysis of dose response curves for siRNA6 and shRNA6 revealed that they had the same EC_{50} value ($p > 0.05$) (Figure 4.3 A). In contrast, shRNA19 was found to be more potent (EC_{50} = 7.4 pM) than siRNA19 (EC_{50} = 93.4 pM) (Figure 4.3 C). Time point (1) RACE-Seq knockdown assays used the same oligo concentrations as reported by Denise et al., (2014). All concentrations, except for shRNA6, gave a luciferase response of greater than 70% knockdown (Table 4.1). For time point (1) RACE-Seq assays, shRNA6 gave only 59.5% knockdown when 0.5 nmol/l oligo concentration was used. This concentration was doubled for the time point (2) RACE-Seq assays resulting in an increase in knockdown to 80.5% (Table 4.1).

The siRNA6 assay was initially conducted at 40 nmol/l for the time point (1) RACE-Seq assay. However, dose response analysis indicated that concentrations from 50 to 5 nmol/l generated a similar response, (Figure 4.3 A) therefore, the concentration of siRNA19 analogue was halved for the RACE-Seq assay at time point (2). Reducing the concentration did not have an effect on luciferase knockdown response with 99.2% knockdown being achieved for the lower concentration (Table 4.1). MTT results for dose response assays indicated no cytotoxicity effects for siRNA6, shRNA6, siRNA19 and shRNA19 for the concentrations used (Figure 4.3 B and D). MTT analysis also confirmed no cytotoxic effects for the individual knockdown assays used for RACE-Seq (Figure 4.3).

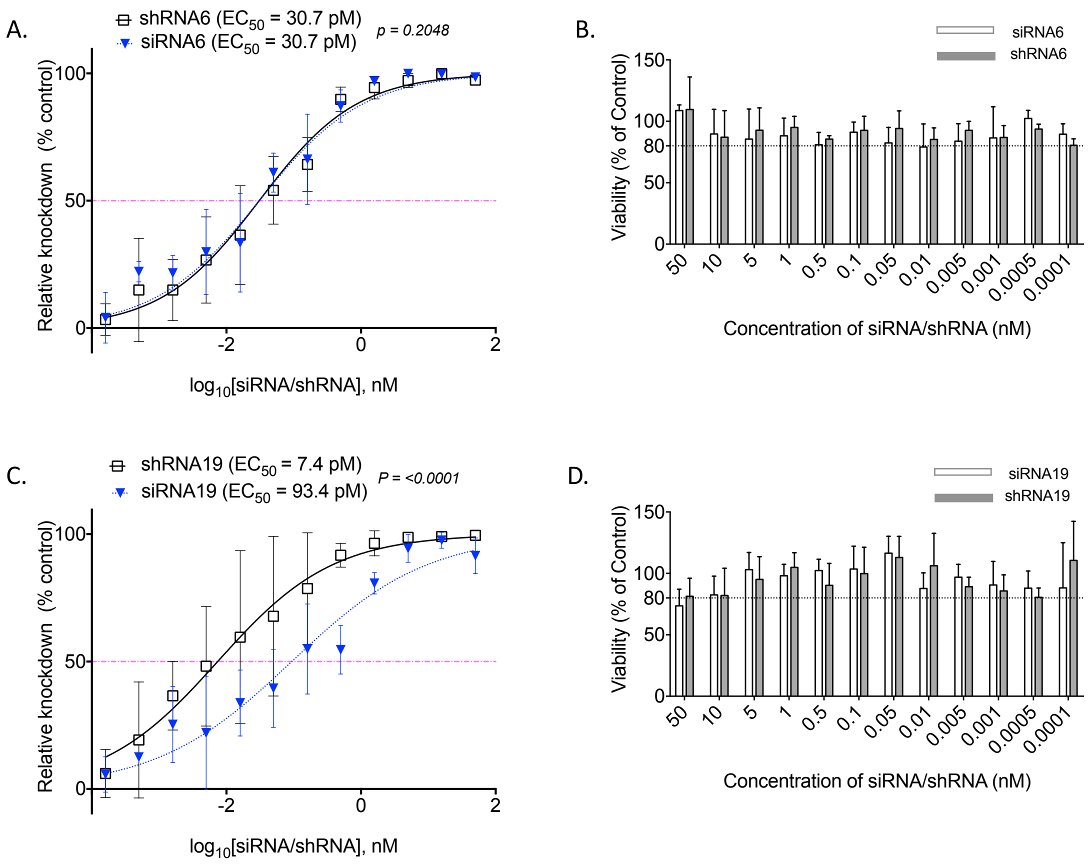


Figure 4.3 Dose response analysis for siRNA6 vs shRNA6 and siRNA19 vs shRNA19. Concentrations for siRNA and shRNA ranged for 50 nM to 0.1 pM for dose response assays for the dose response assays and the MTT viability assays. (A) Dose response curves comparing half maximal effective concentration (EC$_{50}$) indicate siRNA6 and shRNA6 had similar potency (A), while shRNA19 was found to be more potent than siRNA19 (C). Cell viability as determined by the MTT assay did not indicate >25% toxicity for any of the oligonucleotide concentrations used for siRNA6/shRNA6 (B) or siRNA19/shRNA19 (D). The mean and standard deviation of n=6 measurements are shown of two independent experiments with siRNA6/shRNA6 EC$_{50}$= 30.7 pM (p=<=0.2048, siRNA19 EC$_{50}$= 93.4 pM and shRNA19 EC = 7.4 pM (p=<0.0001), Non-linear regression fit.

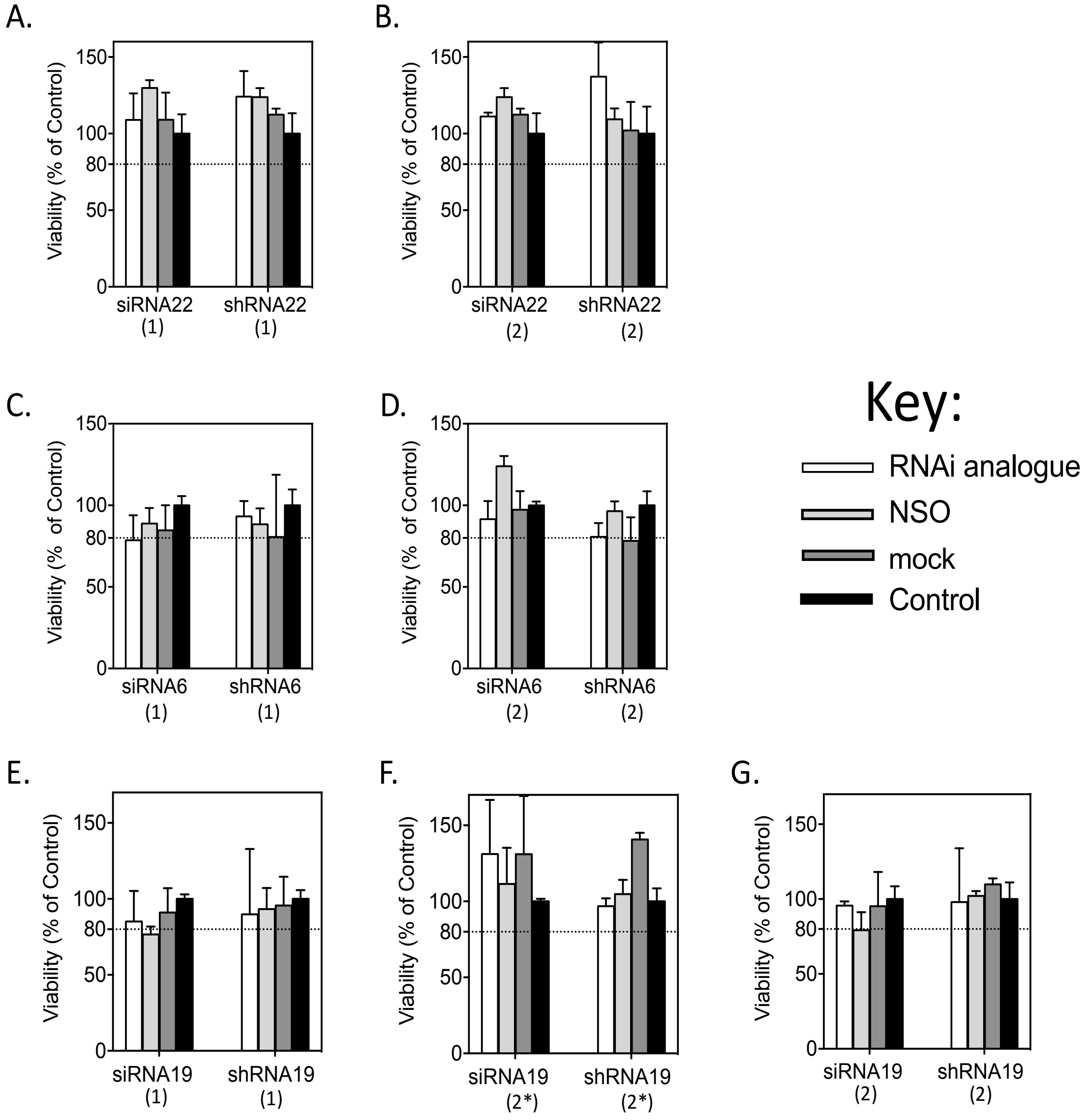


Figure 4.4 Viability assay for RACE-Seq knockdown experiments.

Cell viability as determined by the MTT assay did not indicate >25% toxicity for any of the oligonucleotide. The mean and standard deviation of n=3 measurements are shown for each assay. (NSO = non-specific oligonucleotide, mock = only transfection reagent (Dharmafect3in Opti-MEM added to cells. Control = only Opti-MEM added to cells). The siRNA19* and shRNA19* RACE-Seq samples were processed in a similar manner to the time point (1) samples.

Table 4.1 Concentration of RNAi analogue vs luciferase knockdown achieved.

RNAi Treatment	Time point	Active concentration	Luciferase inhibition (%)	Significance vs NSO control
siRNA22	1	0.5 nmol/l	94.8 ± 0.5	*P*<0.05
siRNA22	2	0.5 nmol/l	97.0 ± 0.3	*P*<0.01
shRNA22	1	0.5 nmol/l	99.3 ± 0.5	*P*<0.05
shRNA22	2	0.5 nmol/l	95.6 ± 0.8	*P*<0.01
siRNA6	1	40 nmol/l	96.5 ± 0.8	*P*<0.05
siRNA6	2	20 nmol/l	99.2 ± 0.3	*P*<0.05
shRNA6	1	0.5 nmol/l	59.5 ± 10	*P*<0.05
shRNA6	2	1.0 nmol/l	80.5 ± 4.6	*P*<0.05
siRNA19	1	10 nmol/l	93.2 ± 2.8	*P*<0.05
siRNA19	2*	10 nmol/l	87.0 ± 4.1	*P*<0.05
siRNA19	2	10 nmol/l	75.9 ± 15	ns
shRNA19	1	0.8nmol/l	78.1 ± 4.8	*P*<0.05
shRNA19	2*	0.8 nmol/l	75.3 ± 8.4	*P*<0.05
shRNA19	2	0.8 nmol/l	77.6 ± 14	ns

The percent luciferase inhibition was determined relative to the non-specific oligonucleotide (NSO) control treatment using the 2-tailes Paired t test. Mean luciferase inhibition and standard deviation from n=3 measurements shown. 2* indicates that the sample processed in a similar manner to the time point (1) samples.

4.2.2 Long amplicon RACE-Seq

4.2.2.1 Impact of RNA input amount on 5' RACE amplicon generation

The first step in the RLM RACE protocol is to tag the 5' end of RISC cleaved RNA with the known RNA adapter sequence using T4 RNA ligase. Total RNA was extracted from treated cells using either the PureLink RNA extraction kit or by the TRIzol method, as indicated in Table 4.2. Both methods produced good quality RNA with A_{260}/A_{280} values 1.7-2.1. However, the extractions using the PureLink kit gave variable A_{260}/A_{230} readings while the TRIzol method was more consistent, giving values of A_{260}/A_{230} = 1.9-2.1. Time point (1) assays used an RNA input amount of 500-600 ng RNA. Various primers were tested before the most suitable reverse transcription and amplification primer combination for the long amplicon RACE-Seq assays was obtained.

The first round PCR for samples siRNA22(1) and shRNA22(1) at time point (1) produced sufficient 5’ RACE product for a single band at 126 bp to be visualised on a 2% agarose gel (Figure 4.5 A). However, siRNA6(1), shRNA6(1), siRNA19(1) and shRNA19(1) samples did not produce visible bands after a single round amplification. For example, Figure 4.5 (C) shows the first round PCR amplification product for siRNA6 and shRNA6 (indicated by * on the gel image). The band at ~500 bp corresponds to the 5’ start of the HCV replicon genome as the site 6 RISC target region lies within the 5’ UTR. This confirmed that at least for site 6, the expected fragment was likely to have been amplified, but at an insufficient amount to be visualised on the gel. To further enrich for the expected fragment, a second round of PCR was carried out. The first round PCR reactions were carried out in a small volume of 15 µl. Then, up to 5 µl was used in the second round PCR reaction. This criterion was specifically designed to minimise selectivity of fragments that might occur when the first round PCR product is diluted or only a small amount is added to the second round PCR. Gel analysis revealed a distinct band at 144 bp for siRNA6(1) but a less distinct band for shRNA6(1) with second round amplification. A second amplified band was also visible just above the expected 5’ RACE band for siRNA6(1) (Figure 4.5 C).

Samples siRNA19(1) and shRNA19(1) had less distinct 5’ RACE bands after second round PCR (Figure 4.5 E). Bands were gel extracted and sent for Sanger sequencing. Although the expected cleavage site could not be determined from this sequencing result, Sanger sequencing confirmed that the expected fragment had indeed been amplified. The correct band/smear could then be identified on the gel and extracted for NGS library preparation.

The time point (2) RACE-Seq assays had a higher amount of RNA (> 1µg) input to the RNA adapter ligation reaction (Table 4.2). The amounts varied depending on the RNA concentration of each individual sample. In order to allow more RNA to be input to the RNA adapter ligation reaction, the volume of the reaction was increased from a 10 µl reaction to a 20µl reaction. This allowed for up to 12.5 µl of RNA to be used in each adapter ligation reaction. Qubit quantification was always used to quantify the RNA for input to the adapter ligation reaction to ensure as accurate a determination for input to the RNA ligation reaction. As previously, siRNA22(2) and shRNA22(2) cleaved RNA samples consistently produced a prominent band at the expected size on 2% agarose gels with the non-transfected control failing to produce a band at the expected 5' RACE position (Figure 4.5 B). Increasing the amount of RNA to the RNA adapter reaction for siRNA6(2), shRNA6(2), siRNA19(2) and shRNA19(2) enabled a single round of PCR amplification to generate a visible band on 2% agarose gels. The siRNA6(2) band was consistently more distinct compared to shRNA6(2) (Figure 4.5 D). This did not impact visualising the bands for excision from agarose gels as the gel extraction application used 4X the amount of and under the blue light transilluminator, bands were suitably visible.

Increasing the amount of RNA input to the adapter ligation reaction enabled the siRNA19(2), siRNA19(2*), shRNA19(2) and shRNA19(2*) 5' RACE amplicons to be visualised after a single round of PCR. The siRNA bands were generally observed at an equal intensity to the shRNA bands for each sample (Figure 4.5 F and G) but a second prominent band was observed just above the expected 5' RACE band for some samples (Figure 4.5 F). The non-transfected control failed to generate a 5' RACE band of the expected size (Figure 4.5 E, F, G).

Table 4.2 Assessment of RNA for 5' RML-RACE

RNAi analogue	Time point	RNA input amount	A_{260}/A_{280}	A_{260}/A_{230}	RNA Extraction method
siRNA22	1	~600 ng	2.06	1.7	Kit
siRNA22	2	1496 ng	2.14	0.65	Kit
shRNA22	1	~600 ng	2.05	1.89	Kit
shRNA22	2	1360 ng	2.15	0.49	Kit
siRNA6	1	~500 ng	2.07	1.66	Kit
siRNA6	2	1355 ng	1.78	1.93	Trizol
shRNA6	1	~500 ng	2.08	1.36	Kit
shRNA6	2	1306 ng	1.78	2.13	Trizol
siRNA19	1	~600 ng	-	-	Kit
siRNA19	2*	1520 ng	2.09	-1.70	Kit
siRNA19	2	1633 ng	1.78	2.12	Trizol
shRNA19	1	~500 ng	-	-	Kit
shRNA19	2*	737 ng	2.07	-1.26	Kit
shRNA19	2	1790 ng	1.81	2.06	Trizol

RNA was extracted from Huh7/con1b cells after treatment with siRNA or shRNA using either the PureLink RNA Extraction Kit (Thermo Fisher Scientific) or the TRIzol method. Both extraction methods generated good quality RNA, but samples extracted using the TRIzol method showed a better A_{260}/A_{280} value. 2* indicates that the sample processed in a similar manner to the time point (1) samples.

Table 4.3 Quantification of DNA input to RACE-Seq library preparation

Time point 1		Time point 2	
Sample	DNA amount (ng)	Sample	DNA amount (ng)
siRNA22	366	siRNA22	196
shRNA22	245	shRNA22	374
siRNA6	301	siRNA6	408
shRNA6	100	shRNA6	248
siRNA19	88	siRNA19*	109
shRNA19	46	shRNA19*	41
		siRNA19	480
		shRNA19	540

The above is the total amount of DNA used at the start of the NGS library preparation for the time point (1) and the time point (2) RACE-Seq assays. The * indicates that the sample processed in a similar manner to the time point (1) samples.

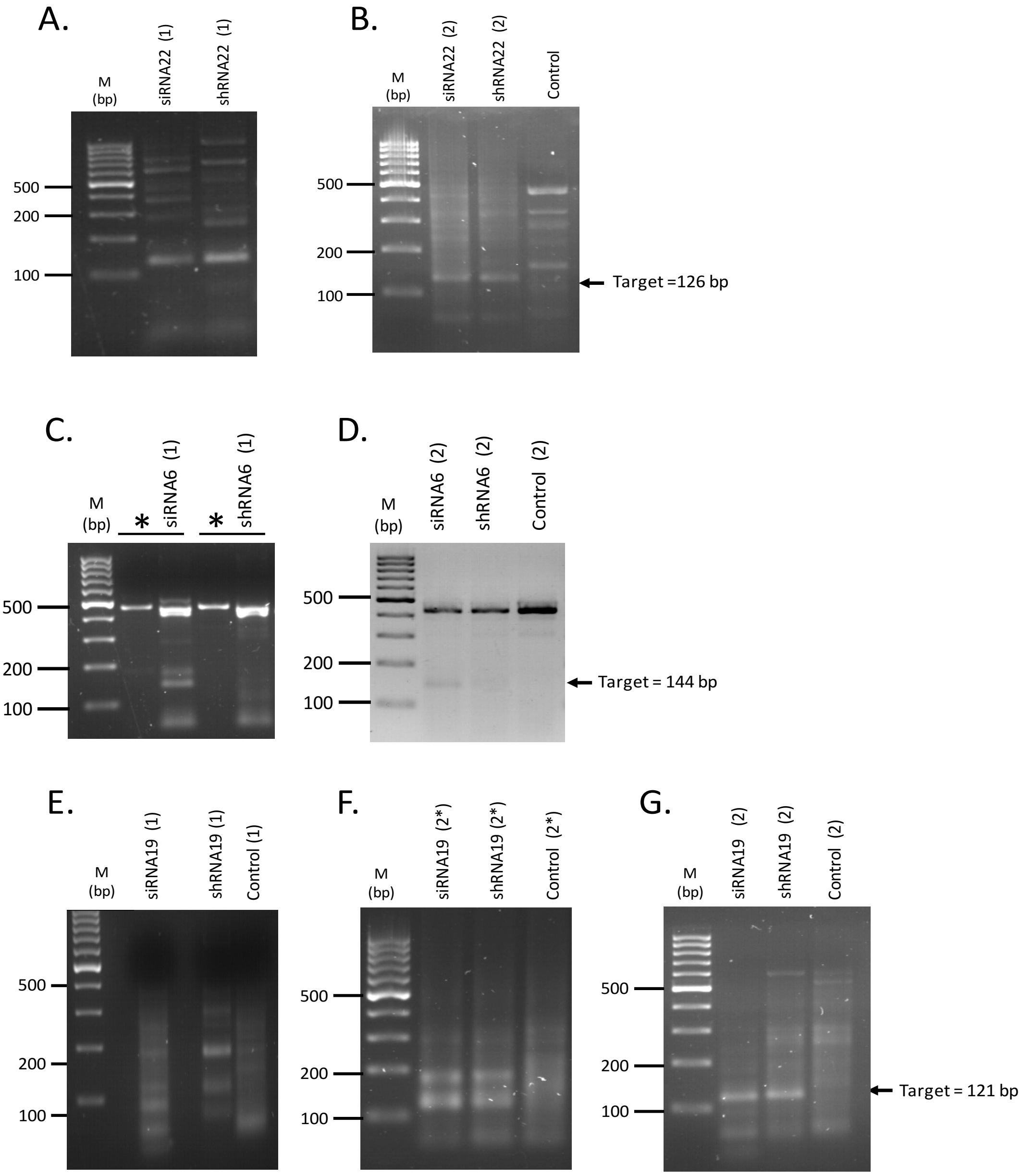


Figure 4.5 Agarose gel analysis of all RLM RACE samples.

Following knockdown of HCV replicon genome, 5’ RACE assays were carried out. HCV positive strand RNA was reverse transcribed with a primer designed to be specific for each of the three RNAi target sites on the HCV genome (site 6, site 19 and site 22). Amplification of the cDNA template of each sample used a gene specific reverse primer internal to the reverse transcription primer and a forward primer to the adapter sequence. All the 5’ RACE PCR reactions were visualised on 2% agarose gels. Assays (A), (C) and (E) were conducted at time point 1, and used up to 600 ng input RNA to the adapter ligation reaction. Samples (C) and (E) required a second round of PCR using the first round PCR product as template in order to visualise the bands on the gel. The (*) in (C) is the first round PCR product. Samples (B), (D), (F) and (G) were the time point 2 samples that used >1 µg input of RNA into the adapter ligation reaction. Samples (F) were gel extracted prior to library preparation while the other time point 2 samples were size selected post library preparation. The size of the expected RACE-PCR band is indicated. M=100 bp GeneRuler molecular ladder, bp = base pairs.

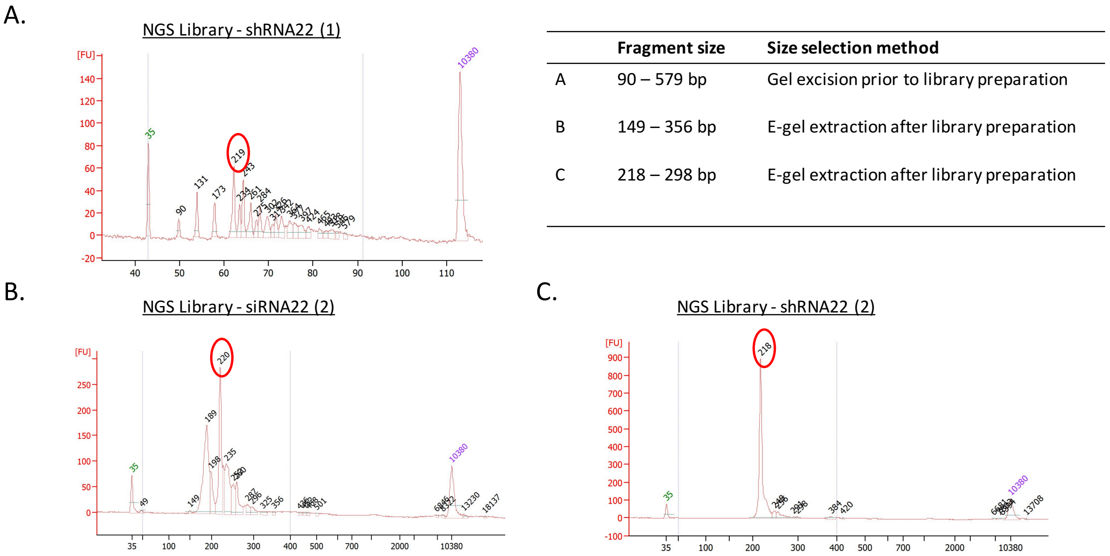
4.2.2.2 Size selection strategies for RACE-Seq assays

Two different strategies were employed for size-selecting the RACE-Seq libraries (Figure 4.2). All time point (1) samples as well as sample siRNA19(2*) and sample shRNA19(2*) were size selected by cutting the DNA bands from a 2% agarose gel prior to library preparation. The purified DNA was quantified by the Qubit DNA assay (Table 4.3) and samples then progressed directly to library preparation. As a consequence, some samples had low (40-110 ng) amounts of DNA available for library preparation after gel extraction (Table 4.3). In order to overcome this limitation, for time point (2) samples, the 5' RACE PCR samples were progressed directly to NGS library preparation without size selection. This allowed for a greater amount of purified DNA to be input to the library preparation (Table 4.3).

Since all NGS libraries were prepared in-house, this allowed the library preparation protocol to be adapted to accommodate the variable DNA input to library preparation. The time point 1 samples used 9 cycles of PCR when enriching for the correct adapter ligated libraries. In contrast, performing the size selection after library preparation allowed greater than 200 ng of purified DNA to be input to NGS library preparation (Table 4.3) and only 6 cycles of PCR was used to enrich for the adapter ligated library. Both strategies generated sufficient amount of NGS library to progress samples to the emulsion PCR stage.

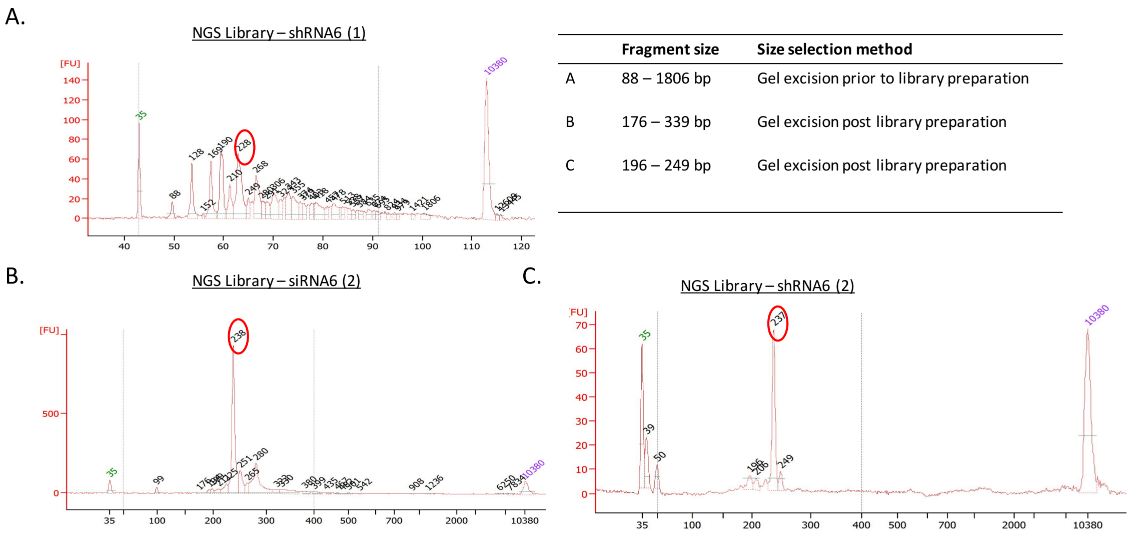
On completion of NGS library preparation, the fragment size range and concentration of samples was assessed on the Bioanalyzer DNA chip. Interestingly, two distinct patterns regarding fragment size range emerged, depending on the method used for preparing the RACE-Seq libraries. Libraries prepared by cutting out the selected band from agarose gels before library preparation generally had a fragment profile that included smaller size fragments (<150 bp) (Figure 4.6 A, Figure 4.7 A and Figure 4.8 A, B and C). On the other hand, RACE PCR samples that were progressed directly to NGS library preparation without prior size selection, lacked these shorter size fragments (Figure 4.6 B and C, Figure 4.7 B and C and Figure 4.8 D and E). The removal of smaller fragments is likely to have been achieved by a combination of bead clean cleanup, which would have eliminated residual primers and short amplicons, and the separation and elimination of smaller fragments by gel electrophoresis. An interesting observation was that, irrespective of size selection strategy, larger fragments (>300 bp) was observed for the 2% agarose gel extractions. This is likely to have occurred due to varied fragment size migration together with the lower precision of slicing agarose gel fractions. In contrast, the siRNA22(2) and shRNA22(2) samples that were size selected by E-gel did not have fragments greater than 300 bp (Figure 4.6 B and C).

In summary, all the samples produced a peak at the expected size for NGS libraries, regardless of the strategy for size selection. Although variability in the fragment size range was observed, the samples were suitable for progressing to emulsion PCR and sequencing.



	Fragment size	Size selection method
A	90 – 579 bp	Gel excision prior to library preparation
B	149 – 356 bp	E-gel extraction after library preparation
C	218 – 298 bp	E-gel extraction after library preparation

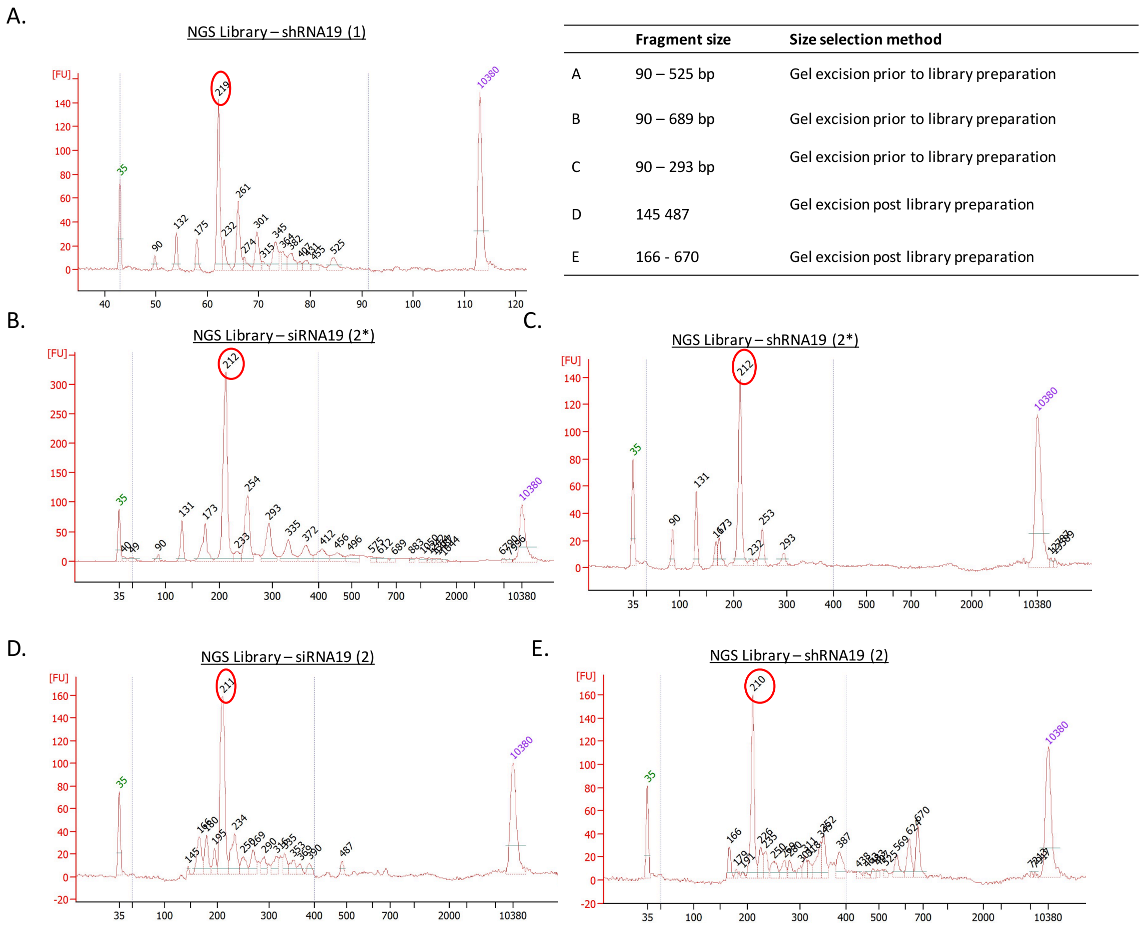
Figure 4.6 Fragment analysis for siRNA22 and shRNA22 RACE-Seq samples. Samples that were size extracted from 2% agarose gels showed a fragment profile that included fragments <150 bp and fragments >300 bp (A). E-gel size selection of siRNA22(2) library sample (B) and shRNA22(2) library sample (C) had fragment profiles that were less dispersed. The expected RACE-Seq peak was readily identified (circled) when analysed on the Bioanalyzer High Sensitivity DNA chip.



	Fragment size	Size selection method
A	88 – 1806 bp	Gel excision prior to library preparation
B	176 – 339 bp	Gel excision post library preparation
C	196 – 249 bp	Gel excision post library preparation

Figure 4.7 Fragment analysis for siRNA6 and shRNA6 RACE-Seq samples.

RACE-Seq library samples that were size extracted from 2% agarose gels showed a fragment profile that included fragments <150 bp and fragments >300 bp (A). For siRNA6(2) library sample (A) and shRNA6(2) library sample (B) the fragment profiles from 2% agarose gel extraction post library preparation had distinct fragment profiles around the expected size range. The expected RACE-Seq peak was readily identified (circled) when analysed on the Bioanalyzer High Sensitivity DNA chip.



	Fragment size	Size selection method
A	90 – 525 bp	Gel excision prior to library preparation
B	90 – 689 bp	Gel excision prior to library preparation
C	90 – 293 bp	Gel excision prior to library preparation
D	145 487	Gel excision post library preparation
E	166 - 670	Gel excision post library preparation

Figure 4.8 Fragment analysis for siRNA19 and shRNA19 RACE-Seq samples.

Samples that were size extracted from 2% agarose gels showed a fragment profile that included fragments <150 bp and/or fragments >300 bp (A, B and C). Additionally, size extraction of libraries from 2% agarose gels post library preparation also produced fragment profiles with fragments >300 bp. The expected RACE-Seq peak was readily identified (circled) for all samples when analysed on the Bioanalyzer High Sensitivity DNA chip. The siRNA19(2)* and shRNA19(2)* samples were processed in a similar manner at time point (1) samples.

4.2.3 Multiplex sequencing on the Ion Torrent PGM

4.2.3.1 Impact of pooled samples on chip loading output

Applying the templated sequencing beads to the sequencing chip is the final physical application stage that can influence the data output. The ISP concentrate is flowed over the chip, and beads deposited into wells by centrifugation. The ISP density image for some RACE-Seq preparations loaded onto the sequencing chip seemed to exhibit a rarely observed phenomenon called bead clumping. This was particularly evident for the shRNA22 legacy sample preparation (Figure 4.9 B) but was also observed for the siRNA22 legacy preparation (Figure 4.9A) where a single RACE-Seq sample was sequenced on each chip. The time point (2) sequencing run also showed evidence of bead climping (Figure 4.9 D). Bead clumping was observed as unevenness of the red heat map pattern on ISP density images. The time point 1 preparation was the only preparation to show greater than 70% chip loading and exhibited an even ISP density profile (Figure 4.9 C). Inhibition zones are also easily identified on three of the ISP density maps as dark blue zones on the periphery of the chip loading area. These zones constitute further loss of data due to failed detection of wells during sequencing and is most often attributed to trapped air bubbles that occur during the chip loading procedure. The chip loading process remains challenging as it is difficult to assess prior to placing the chip on the instrument for sequencing whether any problems at ISP loading has occurred.

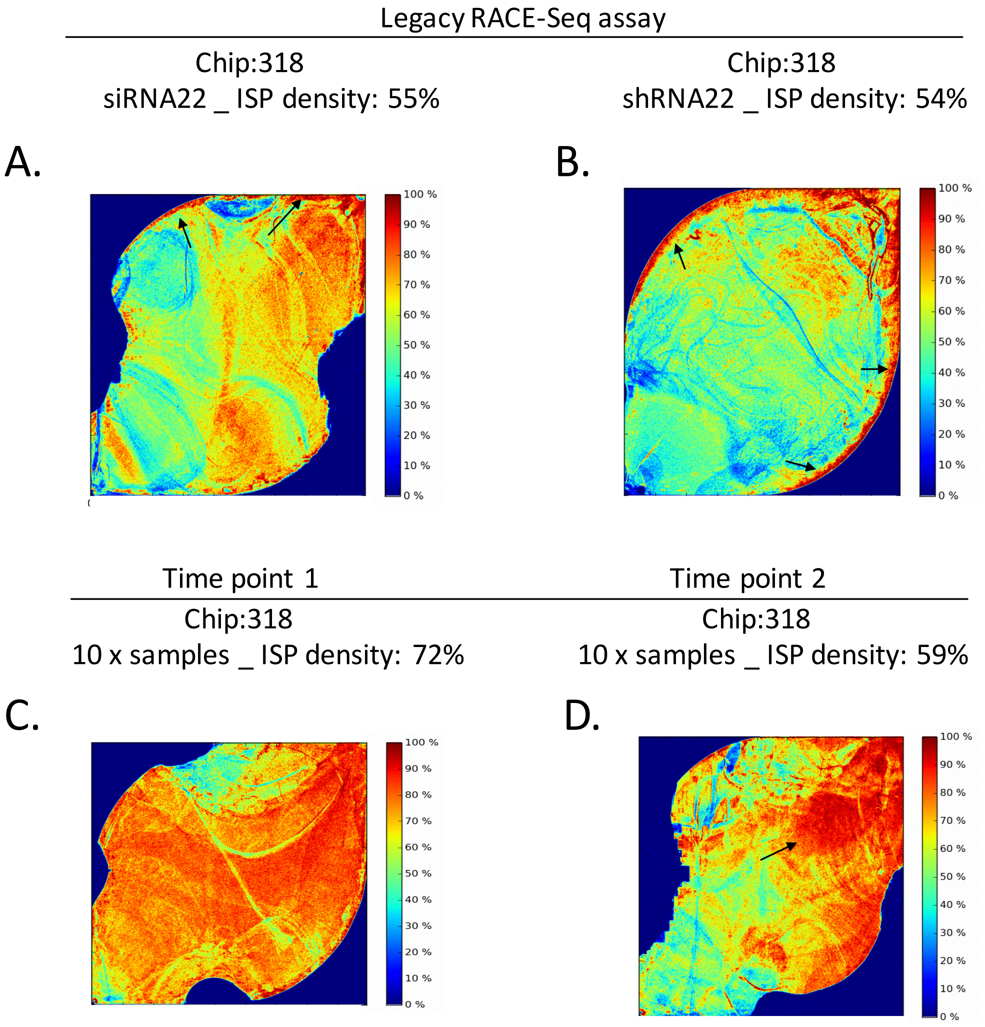


Figure 4.9 ISP density map for Ion Torrent PGM sequencing chips.

Chip loading/ISP varied for the different sequencing runs of the PGM. The legacy RACE-Seq samples had less even spread of ISP deposit across the chip (A) and bead clumping was particularly evident on (B) (arrows) where ISPs had disproportionally settled at the outer region of the chip. The long length pooled RACE-Seq samples showed more even coverage of ISPs across the chip (C) but also had evidence of bead clumping (D).

4.2.3.2 Impact of multiplexing on total number of usable reads obtained per 318 V2 chip

The Ion Torrent Server handles all the data for the PGM. The raw data is interpreted and the base calls determined by proprietary algorithms and a summary of the data is produced. The chip loading percent is related to the number of addressable wells, ie. the number of wells where a sequencing bead was detected. Interestingly, irrespective of chip loading density, all the RACE-Seq preparations had a final usable read count of 2.95 million ± 8% (Table 4.4). Thus, for this set of RACE-Seq samples and the assay conditions used in this study, a threshold level seems to have been reached. The manufacturer threshold level is around 50% of addressed wells, so this falls within the expected minimal output for sequencing on the Ion PGM. In all cases, polyclonal and low quality read filtering reduced the total number of usable reads. For the time point 1 preparation that achieved >8 million deposited ISPs, a higher amount of low quality filtered reads compared to the other sequencing runs contributed to reducing the final read count to below 3 million reads. Since low quality filtering is more likely to relate to the individual samples, and the time point (1) samples were the samples that were extracted from 2% agarose gels prior to library preparation this higher low-quality filter may indicate a negative impact associated with extraction of RACE amplicons from agarose gels prior to library preparation. However, this remains speculation as it is only a single observation. Although all the runs produced sufficient data for downstream RACE-Seq analysis, these preliminary results suggest that for sequencing on the Ion Torrent PGM, loss of read output due to polyclonal and low quality read filtering are key areas for optimisation.

Table 4.4 Ion Torrent PGM data output for RACE-Seq samples

Assay & Chip size	Addressable wells	Chip Loading (%)	Enrichment (%)	Polyclonal (%)	Low quality (%)	Adapter dimer (%)	Usable reads
Legacy _siRNA22	6,191,146	55	99	33	22	4	2,982,359
Legacy_shRNA22	6,046,040	54	99	34	15	4	3,178,404
Time point 1	8,093,181	72	100	35	36	9	2,853,761
Time point 2	6,714,882	59	100	46	17	5	2,766,699

4.3 Discussion

One of the main limitations to the uptake of RACE-Seq for confirming RISC-mediated cleavage events is the lack of reproducible and reliable methods for sample preparation, validation and analysis. Within all the current publications of RACE-Seq, issues relating to sample preparation and data analysis can be identified.

In the Tabernero et al. (2013) publication, they seem to have prepared more than 60 5'RACE samples (the exact number is undisclosed) using the GeneRacer kit protocol. It is unclear which RNA purification strategy was employed after adapter ligation. After reverse transcription and first-round PCR, a second-round PCR was employed to attach the Illumina sequencing adapters to samples and the samples sequenced on an Illumina platform (the platform was not disclosed in the publication). Since no selection for the expected cleavage product size prior to sequencing was performed, together with the fact that for Illumina sequencing, the number of flows determines the length, a large fraction of reads was picked up outside of the expected siRNA target site This would have impacted the data analysis. The datasets were filtered by identifying the exact RNA adapter sequence and removing it and then selecting the next 24 bases for each read. These were collapsed and counted to produce a set of short unique enumerated sequences. These were then aligned to the reference sequence using Bowtie, allowing up to 3 mismatches. The data is presented as percentage of aligned reads. They reported that RACE-Seq samples of only two patients had peaks at the expected cleavage site at significantly different levels to the baseline (from pre-treatment biopsies, banked tumour and normal liver tissue). These two patients had 5' peak levels of ~30% at the expected cleavage site) with the third highest report at ~4%. Inspection of the result image of the distribution of 5' ends for each target indicate that at least 70% of reads were outside the RISC hybridization region of interest, thus biasing the calculation of RISC-induced peak observation.

The Denise et al. (2014) RACE-Seq assay employed a very short reverse transcription length and used the same primer from reverse transcription for the PCR reactions. The short length of PCR fragments and lack of size selection prior to NGS library preparation for sequencing on the Illumina HiSeq 2000 instrument would have resulted in loss of data output due to capture and sequencing of adapter dimer fragments. Of the six data sets presented, one sample is over sequenced with >12 million reads (7 logs) at the expected cleavage site while the other three samples are under sequenced with ~1000 reads (3 logs) at the expected cleavage site indicating problems with sample preparation.

The Barve et al. (2015) data sets clearly indicate a number of issues with sample preparation, as the reported 5' RISC cleaved end counts are reported at counts of under 25 reads at the expected cleavage position for the seven patient samples reported. After adapter ligation, RNA samples were purified using spin columns rather than the standard RNA precipitation cleanup protocol. This is likely to have had the main impact on loss of sample. Then, after two cycles of PCR, the samples were pooled and the sequencing run spiked with 50% Phi X control (a ready to use library with sequencing adapters for Illumina) in an attempt to increase diversity and data quality. The samples were sequenced on the Illumina MiSeq. Certainly, the spike-in control would have significantly reduced the final amount of usable data per sample.

The latest published RACE-Seq dataset by Ganesh et al. (2016) was also sequenced on the Illumina MiSeq instrument. As with Tabernero et al. (2013), the reads were filtered for the RNA adapter sequence, which was removed and then the next 15 nucleotides of each read selected to make the new dataset that was aligned to the reference sequence. They presented their data as percentage of aligned reads but as points rather than the traditional bar plots, which allowed all the datasets to be presented in a single plot. Very distinct peaks were observed at the expected cleavage site for each of the samples, with less than 2% of reads at other positions within the RISC hybridization site. For the expected cleavage position, peak incidence is at approximately 14%, 19% and 22% for the various samples, indicating that some of the reads aligned outside of the RISC hybridization site, similar to Tabernero et al. (2014). Further, the fragment analysis image suggests that RLM RACE enriched for at least one other non-target fragment of 300 bp which was also present for two of the three placebo controls.

The insights gained from critically evaluating these published RACE-Seq results was used improve the development of RACE-Seq libraries for sequencing on the Ion Torrent PGM. Size selection of either the RACE-PCR product or the final library reduced the number of reads that would align outside of the target region and increased the opportunity to obtain a higher proportion of relevant reads. Additionally, by increasing the read length of RACE-PCR amplicons to be at least 100 bp, primer dimer could be effectively removed by bead clean up or size selection on agarose gels or the E-gel system and again enhanced the opportunity to sequence only usable reads. While up to 10 samples could be pooled together on a single 318 V2 chip, developing a high throughput protocol for preparing large numbers of samples remains challenging. An alternative protocol to using RNA precipitation for cleaning up the adapter-ligated RNA is required as this remains the bottleneck to processing multiple samples at one time.

This project represents the first instance of developing the RACE-Seq assay for sequencing on the Ion Torrent PGM. The difference in the Ion Torrent Sequencing platform compared to Illumina has directly impacted on the strategies that evolved for RACE-Seq sample preparation. Some of these strategies are likely to prove valuable for RACE-Seq sample preparation for Illumina sequencing. As with the experience in preparing the legacy RACE-Seq samples, being able to sensitively validate the NGS library size range was valuable for assessing whether NGS samples were suitable for progressing to the next step.

5 RACE-SEQ PREDICTS THE MATURE SIRNA SEQUENCE PROCESSED FROM DICER.

5.1 Introduction

Transthyretin (TTR) is a plasma protein predominantly synthesised by the liver. A very small amount is produced in the area of the brain called the choroid plexus and in the retina (Swiecicki et al., 2015). It serves as a transport protein for thyroid hormone and vitamin A (retinol). More than 100 different point mutations have been identified in the *TTR* gene. These mutations result in misfolded proteins and form extracellular protein deposits in various tissues, including the heart and peripheral nervous system and results in a range of disease conditions leading to mortality. The two hereditary types of disease include Familial amyloid polyneuropathy (FAP) and Familial amyloid cardiomyopathy (FAC). FAP is very rare with approximately 10,000 people affected worldwide, but the disease is well characterised with mutation Val30Met being the most common mutation worldwide (Swiecicki et al., 2015). FAC is most common in African American men over the age of 60. Val122lle has been identified as the most common mutation and is found in almost 4 in 100 African Americans. This has huge implications for the current health care system as liver transplantation is currently the primary treatment for transthyretin amyloidosis (ATTR). The third form of ATTR is termed senile systemic amyloidosis and is caused by non-mutated, ‘wild-type’ TTR amyloid deposits which result in stiffening of the heart due to deposits (Ruberg and Berk, 2012). The disease is slow progressing and is likely to be largely underdiagnosed at present.

Currently, organ transplant and treatment of symptoms is the only course of action available to affected patients. Of the two drug options currently being evaluated, Tafamidis has been approved for use in Europe after showing some slight improvement to patients treated very early in disease diagnosis, but has not been approved by the FDA. Diflunsal, a common pain killer prescribed for conditions such as arthritis, has the risk of severe side effects (Sekijima, 2015). A number of antibody therapy options are also currently under investigation (Palha et al., 2002; Hosoi et al., 2016). Two RNA therapeutics, one an antisense oligonucleotide (ASO) (Benson et al., 2006) and the other a RNAi moiety (Coelho et al., 2013) are currently being investigated in clinical trials. IONIS Pharmaceuticals (previously Isis), leads the way in RNA-targeted drug discovery, and recently announced favourable results for a Phase III clinical trial of its FAP targeting ASO candidate, Ionis-TTR$_{Rx}$. Currently, Alnylam also have Phase II and Phase III clinical trials for their formulated siRNA candidates against FAP (Suhr et al., 2015) and at the end of 2017 announced favourable results for their Phase III clinical trial.

Coelho et al. (2013) reported cleavage confirmation of ALN-TTR02 targeting *TTR* mRNA in a clinical trial using a 5' RACE assay, but the position of the siRNA target site pushed the placement of the 5' RACE reverse and amplification primer positions towards the lower 3' end where optimal primer design was less favourable due to low GC% and numerous short poly(A) and poly(T) stretches. Additionally, the RACE-PCR product would be less and 100 bp, which is less ideal for RACE-Seq. A total of 96 clones from 5' RACE assays for 3 patients was evaluated for the expected cleavage product, before and after therapy and for the three patients, 100%, 98% and 40% of post-treatment RACE clones aligned to the predicted cleavage site (Coelho et al., 2013). Direct confirmation of RNAi activity is greatly underreported in clinical and pre-clinical trials. For example, only 3 of the 32 participants in this trial were evaluated. Further development of the RACE-Seq assay, combined with an intuitive data analysis programme remains a challenge to be overcome.

The aim of this chapter was to perform RACE-Seq on an endogenous human transcript and to evaluate the pattern of 5' end peaks obtained. The *TTR* gene was chosen as a RNAi target as it is highly expressed in the liver. A Dicer -substrate siRNA (DsiRNA) and a siRNA were designed to target *TTR* mRNA in Huh7/con1b cells. The IDT RNAi Design Tool (https://eu.idtdna.com/site/order/designtool/index/DSIRNA_PREDESIGN) was used to design the DsiRNA, while the siRNA sequence was chosen using the *S*fold software package (http://sfold.wadsworth.org/cgi-bin/sirna.pl). The respective compounds were selected based on the design criteria and algorithms for each of the different software, resulting in different but overlapping RNAi target sites being selected for each of the compounds.

5.2 Results

5.2.1 Primer efficiency validation

In order to ensure appropriate use of the $\Delta\Delta C_T$ method, the primer amplification efficiency of TTR and an internal control gene, GAPDH was evaluated. PCR efficiency was established by calibration curves plotting average C_T value vs the log of cDNA template amount (Figure 5.1). R^2 values for calibration curves were r^2=0.95 and r^2=0.94 for TTR and GAPDH samples respectively, indicating poor accuracy between replicates. Poor r^2 values are usually attributed to pipetting errors (Svec et al., 2015). The slope values of each graph was calculated and used to compute the primer efficiency (Table 5.1). Using the slope values, efficiency for TTR amplification was determined as 106.7% while GAPDH amplification efficiency was determined to be 95.2% for this set of experiments. Although the two values fell within the acceptable range of 100% ± 10% as recommended by the MIQE (Minimum Information for Publication of Quantitative Real-time PCR Experiments) guideline, the values have a greater than 10% difference at 11. 5% difference, which is less ideal. The amplification cycles for this efficiency experiment was calculated to be 2.06 for TTR and 1.95 for GAPDH when the slope values were input into the qPCR Efficiency Calculator (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/qpcr-efficiency-calculator.html) (Table 5.1).

Oligo d(T) reverse transcription was chosen as the best method to generate the templates for knockdown determination. The qPCR primers are located upstream of the RNAi target site on the *TTR* transcript, thus reverse transcription with a poly(T) primer or a gene specific primer would generate full length templates only for those RNA transcripts that had not been cleaved, while reverse transcription of cleaved transcripts would terminate prematurely at the RNAi target site, resulting in these truncated templates lacking the qPCR amplification region. Use of random hexamer primers risks amplification of stabilized 5' cleaved targets and could lead to compromised knockdown estimation (Holmes et al., 2010). When designing the siRNA-TTR and DsiRNA-TTR analogues against Human *TTR* transcript NM_00371.3, the poly(A) tail of *TTR* was noted to be only

12 nucleotides in length. Since the aim was to use poly(T) primer for reverse transcription rather than random hexamers, there was a concern that the target would be poorly amplified.

Melt curve analysis showed single unique peaks for samples above 5 ng amount for both *TTR* and *GAPDH* amplification (Figure 5.2 B), and no distinct peak in the no-template-control or DNA contamination control. The samples were analysed by electrophoresis on 2% agarose gel with SYBR Safe stain, confirming single PCR bands and no amplification for the no template control and DNA contamination analysis (Figure 5.2 C).

The primer efficiency analysis indicated that there may be some small discrepancy between GAPDH and TTR amplification efficiency using the proposed protocol. However, since the main aim of the work was not to evaluate knockdown efficacy of siRNA-TTR or DsiRNA-TTR, *GAPDH* was evaluated to be a suitable internal control for this set of experiments.

Table 5.1 Reporting of amplification efficiency for TTR and GAPDH

	Slope value	PCR efficiency*	Amplification cycle*
TTR	-3.1846	106.7%	2.06
GAPDH	-3.4425	95.2%	1.95

* Amplification efficiency and amplification cycle were calculated from slope values of calibration curves in Figure 5.1, using Thermo Fisher Scientific QPCR Efficiency Tool (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo -scientific-web-tools/qpcr-efficiency-calculator.html).

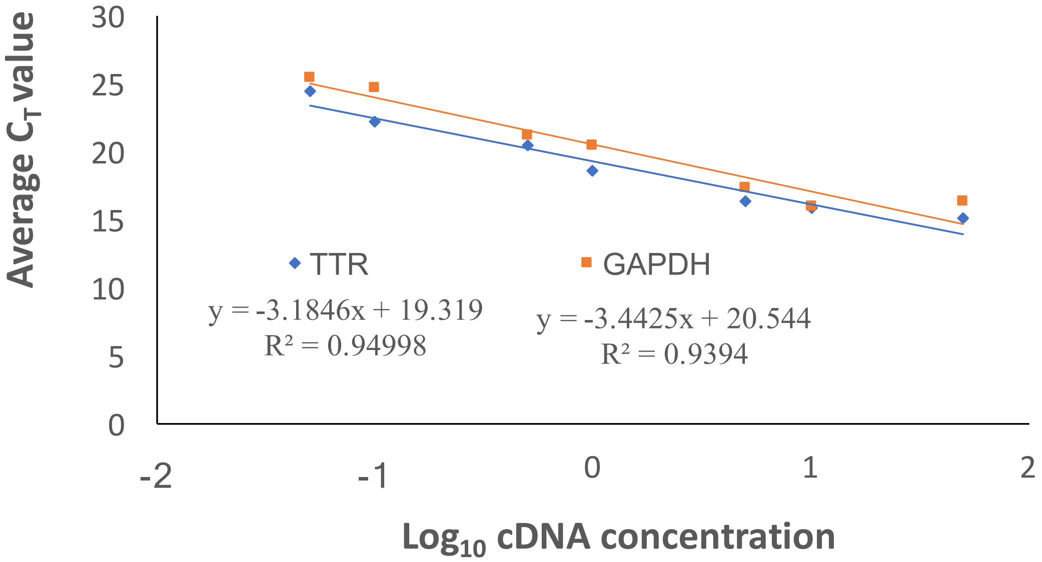


Figure 5.1 Determining amplification efficiency by calibration curve analysis. cDNA was prepared at 7-point dilution (50 nM, 10 nM, 5 nM, 1 nM, 0.5 nM, 0.1nM, 0.05 nM). Points are representative of average C_T value from triplicate wells.

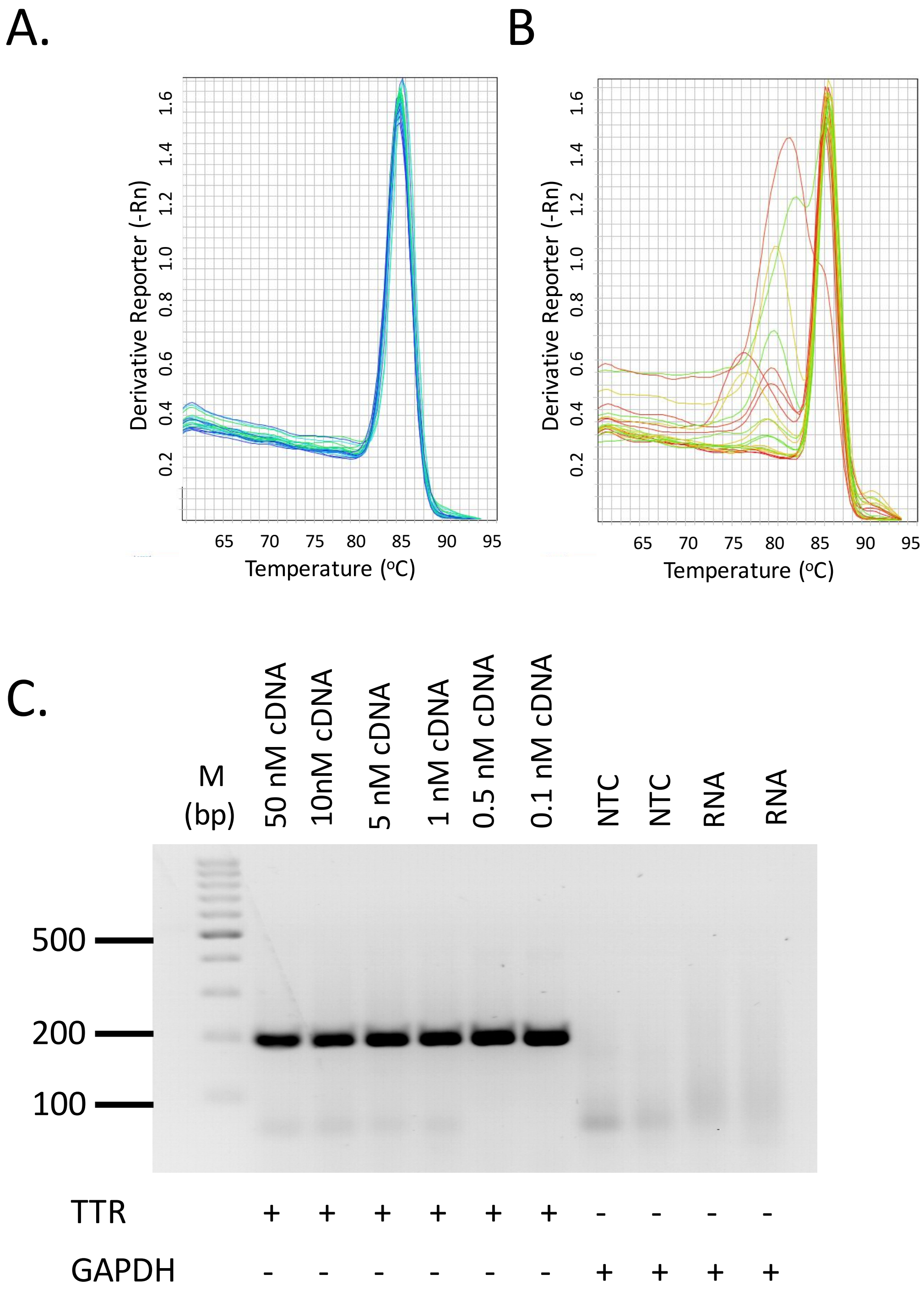


Figure 5.2 Melt curve analysis of PCR efficiency reaction.

Melt curve analysis for GAPDH (A) confirmed a single peak generated for all dilutions of cDNA template. Melt curve analysis of *TTR* products (B) indicates some spurious amplification for concentrations below 5 nM. Agarose gel analysis of *TTR* amplified reactions confirm a single band for all dilutions (C). No template control was positive for no amplification of GAPDH. No DNA contamination in RNA was positive for no amplification of GAPDH when RNA was used as template (C). NTC=no template control; RNA= template for amplification; M=100 bp GeneRuler DNA ladder.

5.2.2 Targeted knockdown of TTR mRNA expression in Huh7/con1b cells

DsiRNA-TTR and siRNA-TTR targeting two different, but overlapping regions on *TTR* transcripts in Huh7/con1b cells were successful in reducing levels of mRNA target (Figure 5.3) with siRNA-TTR achieving 79.5% knockdown and DsiRNA-TTR achieving 75.5% knockdown when each treatment was applied at 10 nM effective concentration.

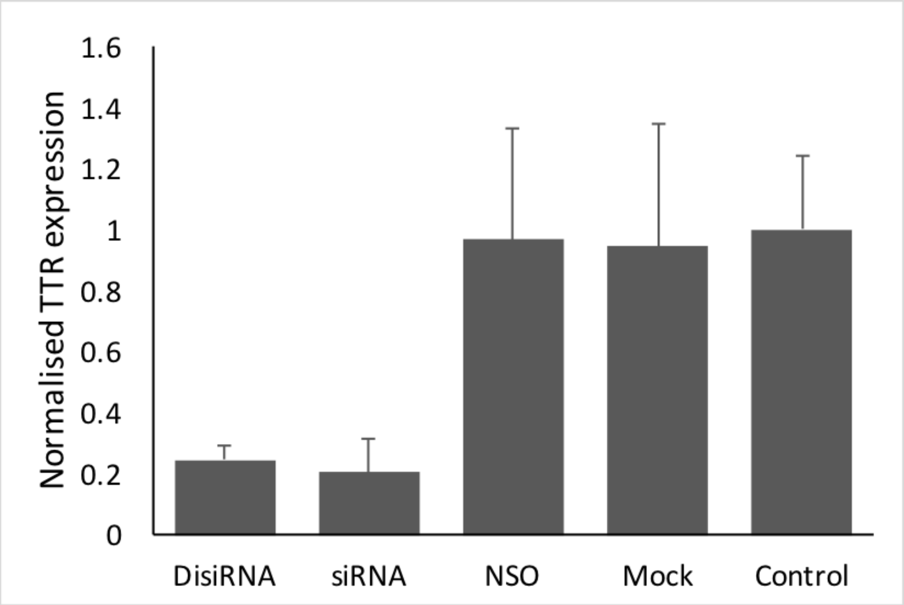


Figure 5.3 Relative *TTR* mRNA levels after siRNA-TTR treatment and DsiRNA-TTR treatment in a hepatocellular cell line, stably expressing HCV replicon RNA.

5.2.3 Validation of RACE-Seq libraries by fragment analysis

In preparing the 5'RACE-PCR samples, 1 µg or greater had already been established to generate 5' RACE bands in a single round of PCR (Chapter 4). RNA adapter ligation was done in a 20 µl volume with the maximum allowed volume of RNA input of 12.5 µl. Care was taken to elute RNA that would be used for 5' RML-RACE into a small volume (~30 µl or less). This helped to ensure that > 1.0 µg RNA was input to adapter ligation reactions (Table 5.2). After adapter ligation and RNA clean up, 5' RLM-RACE cDNA templates were prepared as previously described (Section 2.9) using primer TTR+L1 for reverse transcription. Amplification was carried out using primers GeneRacerF1 and TTR+L2 primers. PCR reactions were as previously described (Section 2.9.1) and 8 µl of PCR sample analysed on a 2% agarose gel with SYBR Safe nucleic acid stain (Figure 5.4 A). Gel analysis identified prominent bands of the expected size, ~170 bp, for both the siRNA-TTR and DsiRNA-TTR samples but not the scrambled control. Additional background amplification was easily visible for all three samples. Of note is that a smear of background amplification was clearly present for the scrambled control sample, but no prominent size band was seen at the expected position for RISC cleavage.

In order to prepare amplicon samples for library preparation, PCR reactions were first cleaned using 1.6X Ampure beads. Samples were quantified and siRNA-TTR and DsiRNA-TTR NGS libraries prepared as previously described (Section 2.10). Since the gel analysis showed very prominent bands for both the samples, E-gel size selection was used to extract the expected NGS band size. NGS fragment size after E-gel extraction was analysed on the Bioanalyzer High Sensitivity chip. For both siRNA-TTR and DsiRNA-TTR sample, E-gel extraction of 5'RACE NGS fragments was generally in the range of 172 bp to 292bp (Figure 5.4 B and C). However, two distinct fragment profiles were seen for each sample. A quite distinct single peak at ~250 bp was isolated for siRNA-TTR (Figure 5.4 B), but the DsiRNA-TTR sample had a cluster of very prominent peaks at 172 bp to 221 bp, as well as the prominent ~ 250 bp peak (Figure 5.4 C). Only a few >300 bp peaks were picked up in Bioanalyzer analysis, but these were at a very low concentration. For both samples, the E-gel isolated NGS libraries met the criteria for sequencing on the Ion Torrent for 200 bp sequencing. For these samples, an observation of discrete 5' RACE-PCR bands correlated well with using the E-gel system for isolating the desired fragments. Isolating the fragments after library preparation very efficiently eliminated <150 bp fragments for both samples. The E-gel system elutes the DNA fragments into wells that are topped up with ~25 µl nuclease-free water and the fragments aspirated at the appropriate point. Since the libraries were prepared in-house, the purity of the final library was not assessed. However, if the samples were to be sent for external sequencing, it is likely that a minimum $OD_{260/280}$ = ~2.0 would be required. Since the DNA passes through a gel matrix, it is feasible that some buffer components transfer into the collection well and interfere with purity determination. The DsiRNA-TTR and siRNA-TTR samples were pooled with the HCV 5’RACE samples, samples quantified, diluted and transferred to emulsion PCR and then progressed to sequencing.

Table 5.2 Nucleic acid quantification for 5’RML-RACE and RACE-Seq

	Qubit Quantification	Nanodrop analysis			Amount RNA input to 5' RLM-RACE	Amount bead cleaned DNA input to NGS
Sample	Average [ng/ul]	Average [ng/ul]	$A_{260/280}$	$A_{260}230$		
DsiRNA-TTR	340	405	2.04	1.57	2,105 ng	270 ng
siRNA-TTR	302	350	2.03	1.39	2,395 ng	282 ng
Dscr	302	363	2.04	1.63	1,428 ng	119 ng

(Total RNA is calculated as Qubit quantified amount RNA x 12.5 µl)

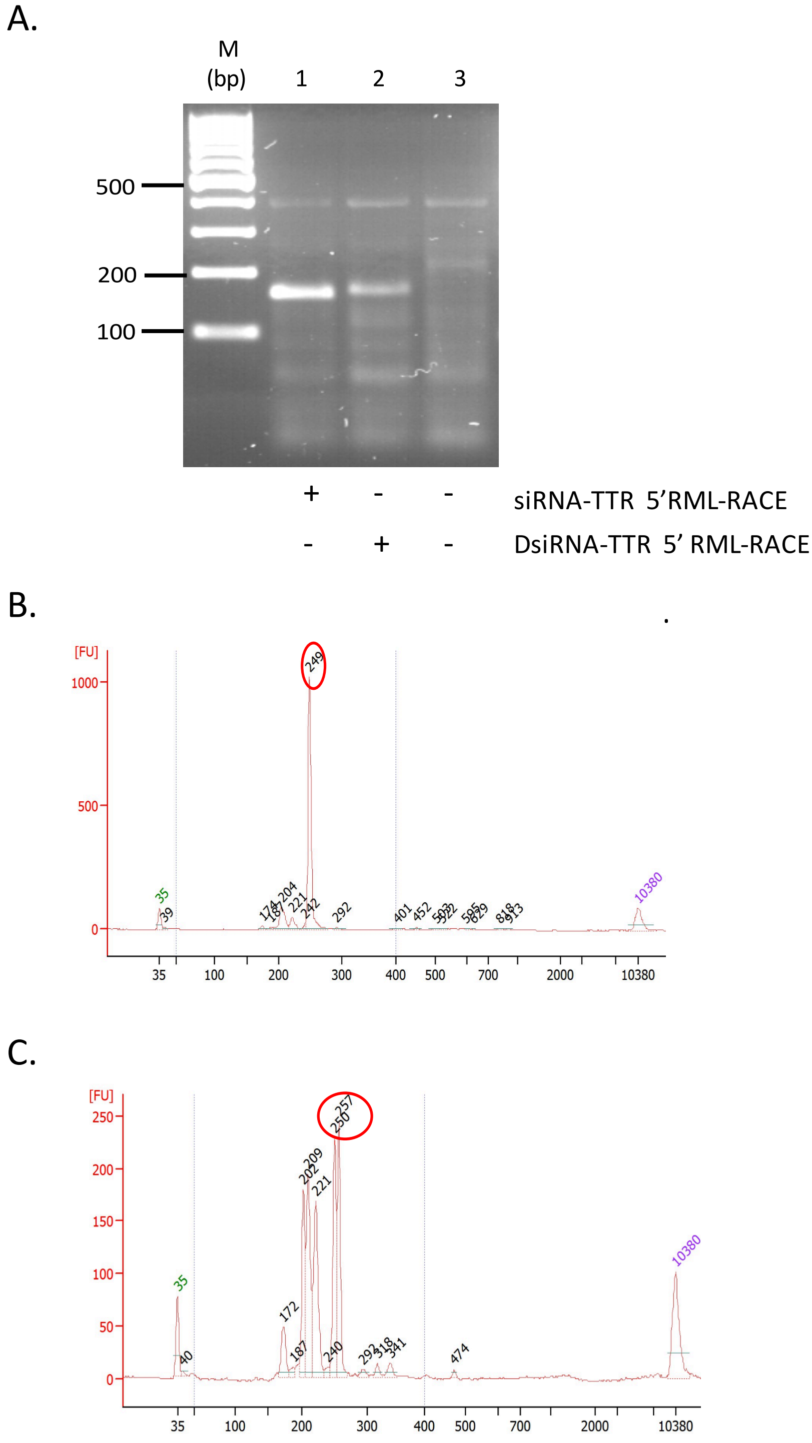


Figure 5.4 Visualisation of 5'RML-RACE amplicons and RACE-Seq libraries post E-gel size selection.

Distinct bands at the expected size (~170 bp) were obtained for 5'RML-RACE siRNA-TTR and DsiRNA-TTR samples (A). NGS libraries were prepared from 5'RML-RACE samples and the expected size band extracted using the E-gel Size Selection System. A discrete fragment range was observed for siRNA-TTR size selected RACE-Seq sample (B), while the DsiRNA-TTR sample had a peak fragment range 172 bp to 257 bp (C).

5.2.4 RACE-Seq data analysis

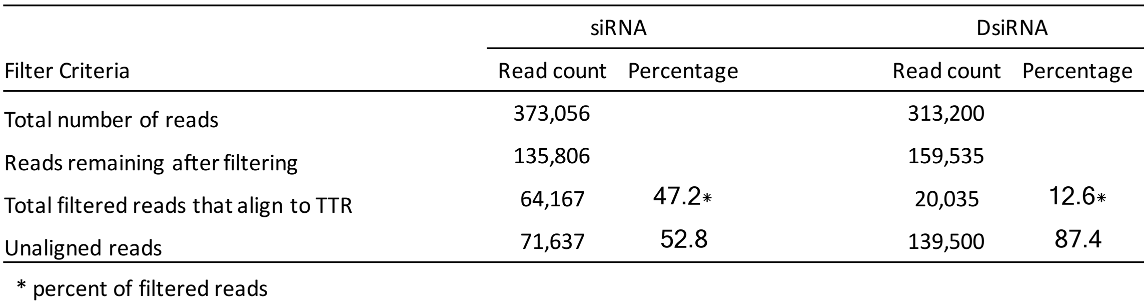
RACE-Seq sample siRNA-TTR, generated 373,056 reads, while RACE-Seq sample DsiRNA-TTR generated 313,200 reads (Table 5.3). After filtering the data and aligning reads to the reference, 64,167 reads and 20,035 reads aligned to the reference for siRNA-TTR and DsiRNA-TTR respectively (Table 5.3).

To identify why the rest of the filtered reads failed to align to the reference, the reads were collapsed and mapped to the *TTR* reference. This identified that for siRNA-TTR, the unaligned data was in fact primarily *TTR* sequence that had failed to align due to mutations, insertions or deletions. A poly-C5 region 17 bases from the expected siRNA-TTR RISC cleavage site seemed to be particularly problematic. For the DsiRNA-TTR sample, mapping collapsed reads to the *TTR* reference revealed non-TTR sequence that was identified as contaminating HCV sequence that aligned to site 19 on the HCV replicon reference sequence. This contamination is likely to have occurred at a stage prior to ligation of sequencing adapters as the contaminating reads had been de-multiplexed to the respective TTR barcoded samples. Interestingly, these contaminating HCV reads seem to have either held the *TTR* amplification primer sequence or escaped the read filtering process, as one of the pre-filtering steps selects for reads containing the TTR+L2 sequence. Therefore, a likely conclusion is that the contamination occurred during amplification of TTR 5’ RACE cDNA.

The loss of aligned data due to sequence variation beyond of the RISC cleavage site and presence of contaminating/non-target sequence led to a new strategy for filtering the RACE-Seq data (Figure 5.5). Since the sequence just after the RNA adapter is the region of interest, a strategy that shortens all the reads to a fixed distance beyond the RNA adapter was a sensible approach and had previously been implemented for RACE-Seq data filtering (Tabernero et al., 2013; Ganesh et al., 2016). A strategy that selects 20 bases after the RNA adapter was chosen due to the homopolymer C region in close proximity to the siRNA-TTR RISC cleavage site. Additionally, the short 20 bases length would also be more suitable to analysing the HCV RACE-Seq data as the HCV replicon sequence was expected to be highly variable. To eliminate contaminating reads and non-specific amplicons from the dataset, a strategy to specifically filter for the target sequence was implemented (Figure 5.5).

The FASTQ datasets were filtered by searching for the RNA adapter sequence in both directions, reverse complementing the reads that were in reverse order and then adding these to the reads with the correctly orientated adapter sequence. The gene specific reads were selected by first removing the RNA adapter and then searching for reads that had the expected 10 bp sequence at the 3' end. This gave a filtered data set of 112,939 reads for siRNA-TTR and 64,257 reads for DsiRNA-TTR. The GeneRacer adapter sequence was then removed and reads aligned to *TTR*. This found that 90,252 reads aligned to TTR for the siRNA-TTR sample and 51,105 reads for DsiRNA-TTR (Table 5.4). Thus, the new data filtering strategy increased the total number of aligned reads from 64,167 to 90,252 reads for siRNA-TTR and from 20,035 to 51,105 reads for DsiRNA-TTR (Table 5.3 vs Table 5.4). For DsiRNA-TTR, the number of filtered reads decreased from 159,535 reads to 64,257 reads, but represented reads that were truly the expected target sequence.

Table 5.3 RACE-Seq read analysis using filtering to retain long length reads



	siRNA		DsiRNA	
Filter Criteria	Read count	Percentage	Read count	Percentage
Total number of reads	373,056		313,200	
Reads remaining after filtering	135,806		159,535	
Total filtered reads that align to TTR	64,167	47.2*	20,035	12.6*
Unaligned reads	71,637	52.8	139,500	87.4

* percent of filtered reads

Table 5.4 Read analysis for RACE-Seq siRNA-TTR and DsiRNA-TTR

	siRNA		DsiRNA	
Filter Criteria	Read count	Percentage	Read count	Percentage
Total number of reads	373,056		313,200	
Total reads with GeneRacer adapter	171,000	45.8	148,357	47.4
Total reads with inverted GeneRacer sequence	148,448	39.8	116,315	37.1
Total reads with GeneRacer adapter	319,448	85.6	264,672	84.5
Reads with reverse primer sequence(trimmed)	193,259		221,422	
Reads specific to target (10 bp internal)	144,578		77,770	
Total filtered reads (adapter trimmed)	112,939	30.3	64,257	20.5
Total reads that align to TTR	90,252	79.9*	51,105	79.5*
Unaligned reads	22,687	20.1*	12,228	20.5*

* percent of filtered reads

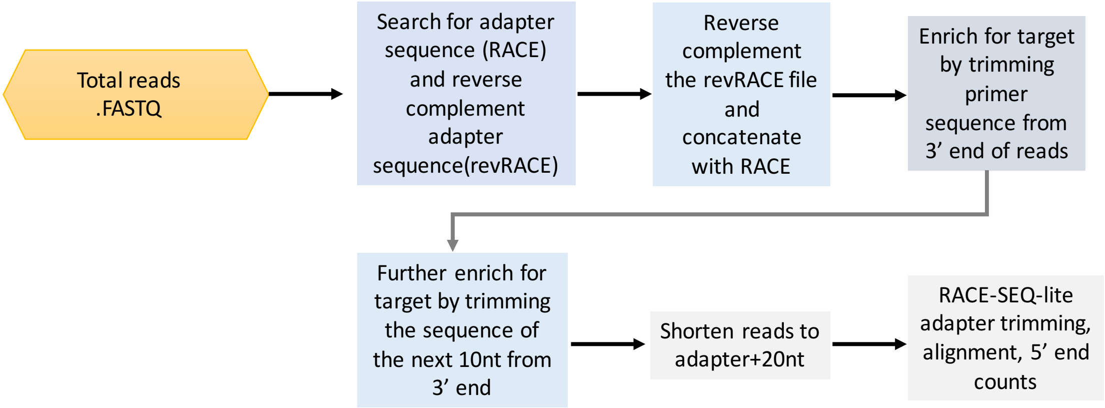

Total reads .FASTQ
Search for adapter sequence (RACE) and reverse complement adapter sequence(revRACE)
Reverse complement the revRACE file and concatenate with RACE
Enrich for target by trimming primer sequence from 3’ end of reads
Further enrich for target by trimming the sequence of the next 10nt from 3’ end
Shorten reads to adapter+20nt
RACE-SEQ-lite adapter trimming, alignment, 5’ end counts

Figure 5.5 Data analysis strategy for filtering RACE-Seq datasets to generate the adapter+20-nucleotide short filtered datasets.

5.2.5 RACE-Seq confirms the expected cleavage product for siRNA-TTR

The siRNA-TTR RNAi bioactive had been designed to be a standard 21 bp duplex siRNA with 3’ 2-nt DNA overhangs (Figure 5.6 A). Aligning the filtered RACE-Seq siRNA-TTR dataset to the *TTR* reference gene resulted in 79.9% of aligned reads with a 5’ end at the expected RISC cleavage position, C=483 (Table 5.5). This corresponded exactly to canonical siRNA-mediated cleavage of the target between bases 10-11 from the 5’ end of the expected guide strand. Interestingly, RACE-Seq reported an additional 5’ end peak at +1 of the expected cleavage site (Figure 5.6 B). Whether this second peak of 5’ ends was generated by guide strand-mediated RISC activity remains unclear. For this peak to have been generated by canonical RISC cleavage, the siRNA would need to have been cleaved such that the guide strand now had nucleotide two as the first nucleotide, but retained 3’ and 5’ structural elements necessary for AGO2 recognition. Confirmation of such a sequence can only be obtained by capturing the siRNA fraction of active RISC by immunoprecipitation of AGO2 and isolating and sequencing the small RNA fraction. The observation of a second 5’ peak at +1 position suggests that RACE-Seq has the potential to provide novel information regarding on-target RISC activity for siRNA duplexes generating multiple guide strand species.

5.2.6 RACE-Seq identifies the sequence of the active guide strand processed out of Dicer for DsiRNA-TTR

Figure 5.7 (A) illustrates the predicted Dicer processing of DsiRNA-TTR to produce a 21-mer or 22-mer siRNA. DsiRNA-TTR was designed to be a 25-27 duplex with a 2-nt 3’ overhang to promote directional processing by Dicer and preferential strand selection. The other end is blunt ended to preclude Dicer binding at that end. Two DNA nucleotides are included at the blunt end of the sense strand. Thus, Dicer recognition of the 3’ overhang enforces asymmetric loading. Dicer processing is expected to result in a major 21-mer and may produce a minor 22-mer active siRNA (Amarzguioui and Rossi, 2008).

When the RACE-Seq data for DsiRNA was aligned to the *TTR* reference, a peak at G=475 emerged, rather than the expected T=474 (Table 5.5). This led to the conclusion that Dicer processing had occurred at 5-nt internal to the guide strand, rather than the expected 6-nt internal to the guide strand (Figure 5.7 A). In conclusion, RACE-Seq analysis could identify the sequence of the guide strand that must have been processed out of Dicer.

5.2.7 RACE-Seq predicts that the unprocessed DsiRNA-TTR is itself a substrate for AGO2-RISC

For the DsiRNA-TTR RACE-Seq sample, a further unexpected observation was noted. A distinct peak at position T=480 of 25.8% aligned reads was generated (Table 5.5). Initially, this peak had been classed as putative degradation product. Surprisingly the second peak exactly corresponded with canonical RISC cleavage of target if the 27-mer RNA of DsiRNA-TTR is used as a guide strand. (Figure 5.7 B). Whether the entire 27-mer guide strand sits within the active RISC or a shortened version of DsiRNA-TTR was derived is unclear. Dicer processing of 25/27 duplex siRNA is expected to yield primarily a 21-mer siRNA (Kim et al., 2005). Canonical Dicer processing is expected to cleave 6-from the 27-mer guide strand, so that, if the opposite end of the duplex were to be cleaved, this would yield a shorter duplex with compatible ends for AGO2 recognition. Such a theory remains to be validated.

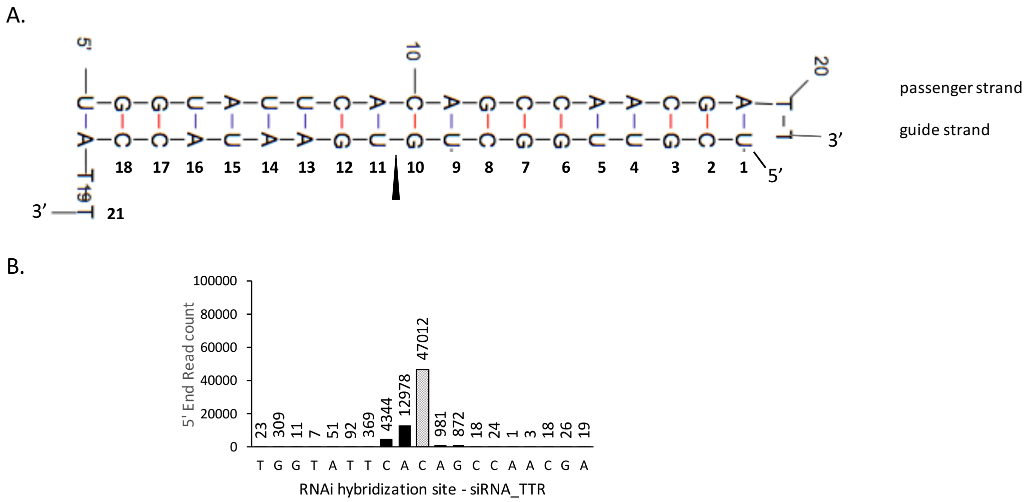


Figure 5.6 RACE-Seq result for siRNA-TTR.

(A) Diagram illustrating the expected cleavage position for siRNA-TTR, a 21-mer oligo with 19 bp duplex stem and 2-nt overhang. (B) RACE-Seq confirms the expected RISC cleavage site at the highest peak. (TT) at the 3' and 5' overhangs = DNA; siRNA-TTR duplex is labelled in 5' to 3' direction of guide strand (target hybridization orientation); grey bar = RISC cleaved site.

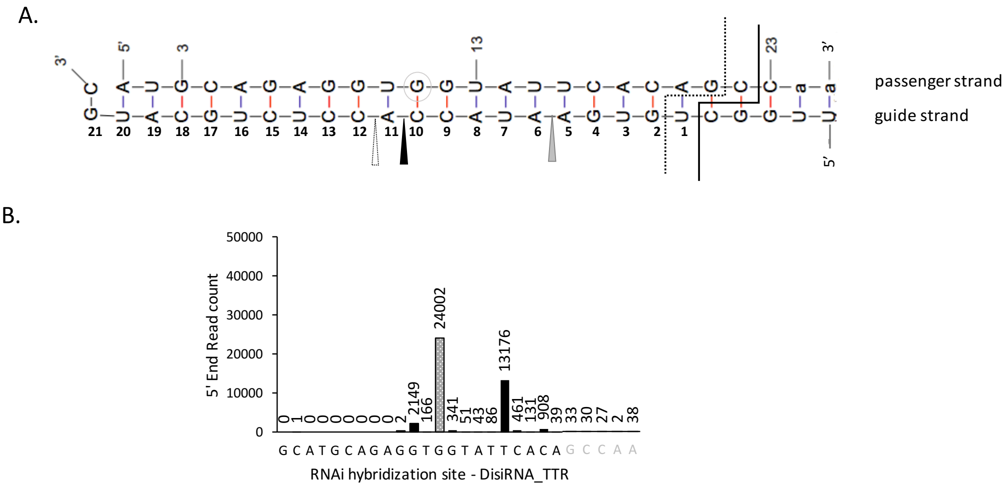


Figure 5.7 RACE-Seq identifies the siRNA guide strand of DsiRNA-TTR

(A) Illustration of the DsiRNA-TTR duplex indicating the expected Dicer cleavage site (dotted line) which would generate a 21-mer duplex and expected cleavage position at open triangle. The RACE-Seq predicted Dicer processed position is indicted by the solid black line, to generate an expected 22-mer product (A). A distinct 5' peak confirms RISC activity of a 22-mer siRNA processed out of Dicer (B). (aa) at 3' end of passenger strand = DNA; DsiRNA duplex is labelled in 5' to 3' direction of guide strand from the expected 22-mer 5' end. grey bar = RISC cleaved site.

Table 5.5 Table indicating the percent aligned reads for siRNA-TTR and DsiRNA-TTR from RACE-Seq assays

siRNA-TTR				DsiRNA-TTR			
siRNA at hybridization site	*TTR* nucleotide position	Nucleotide	Aligned reads (%)	DsiRNA at hybridization site	*TTR* nucleotide position	Nucleotide	Aligne reads (%)
3'	474	T	0.0	3'	463	G	0.0
	475	G	0.8		464	C	0.0
	476	G	0.0		465	A	0.0
	477	T	0.0		466	T	0.0
	478	A	0.0		467	G	0.0
	479	T	0.0		468	C	0.0
	480	T	0.5		469	A	0.0
	481	C	0.0		470	G	0.0
	482	A	11.4		471	A	0.0
	483	C	79.9		472	G	0.0
	484	A	1.2		473	G	4.2
	485	G	0.1		474	T	0.3
	486	C	0.0		475	G	47.0
	487	C	0.0		476	G	0.7
	488	A	0.0		477	T	0.1
	489	A	0.0		478	A	0.1
	490	C	0.0		479	T	0.2
	491	G	0.0		480	T	25.8
5'	492	A	0.0		481	C	0.9
					482	A	0.3
					483	C	1.8
					484	A	0.1
					485	G	0.1
					486	C	0.1
					487	C	0.1
					488	A	0.0
				5'	489	A	0.1

RACE-Seq datasets were pre-filtered to specifically select for the target sequence and then trimmed to adapter+20-nt length. These short reads were input to the RACE-SEQ-lite analysis pipeline which performed the alignment reported the 5’ end counts as a percent of total aligned reads. For siRNA-TTR, position 483 = C = 79.89% corresponds to a count of 10 from the 5’ end of the hybridized siRNA. For DsiRNA-TTR, two major peaks are observed (red); position 480 = T = 25.8% corresponds to a count of 10 from 5’ end of the unprocessed DsiRNA-TTR guide strand and position 475 = g = 47.0 % corresponds to a count of 10 for a proposed Dicer processed guide strand with start position 458 (grey).

5.3 Discussion

The present study was designed to assess the RACE-Seq protocol (implemented in Chapter 4) and accompanying analysis pipeline when applied to a genomic transcript. The *TTR* gene was chosen as the test transcript as it is known to be highly expressed in hepatic cells (Costa et al., 1990). A standard siRNA of 19 bp duplex and 2-nt DNA overhang was designed using a web-based tool. The algorithms utilised in such design software incorporate extensive experimental and computation analysis of validated potent siRNAs (Reynolds et al., 2004). Dicer processing of longer RNAi triggers has been implicated to have higher potency compared to 21-mer triggers as Dicer processing is believed to promote more efficient incorporation into RISC (Ketting et al., 2001; Boudreau et al., 2008; Snead et al., 2013). The design of DsiRNA often begins by choosing an optimal 21-mer and then extending the 5' end by 6 nucleotides and designing a blunt end that terminates in two DNA bases. The free 3' 2-nt overhang on the guide strand is the signal for Dicer capture while the blunt-end is thought to be recognised in a similar manner to shRNA loop structures (Amarzguioui et al., 2006).

In this Project, we present a computationally simple workflow for analysing RACE-Seq data. Being able to implement the adapter removal, sequence enrichment and any quality filtering analysis as a stand-alone operation allowed in-depth analysis of the datasets. While each command-line code was executed manually as a single line, it is possible to write simple code that would execute the desired commands and produce a summary output of text and graphics.

For the siRNA-TTR sample, RACE-Seq reported the expected cleavage site as the major peak, however a second peak at +1, and third peak at +2 were also reported. Whether these additional peaks were generated by siRNA-mediated RISC activity remains unknown, but would require that a fraction of the siRNA-TTR underwent processing *in vivo*, such that the 5' end of the guide strand was cleaved by 1 and 2 bases, but retained the 2'nt overhang and 5' phosphate for AGO2 recognition.

On the question of Dicer processing of the DsiRNA, the expectation was for a predominantly 21-mer active product and minor 22-mer product (Amarzguioui and Rossi, 2008). However, for DsiRNA-TTR, RACE-Seq identified a cleaved target that points to a preference for a 22-mer Dicer processed RNAi trigger from DsiRNA-TTR. A major 22-mer has previously been identified from Dicer processed 25/27-mer by Snead et al., (2013) who also identified a major 22-mer processed out of Dicer for a 25/27-mer and by deep sequencing the small RNAs from DsiRNA transfected cells, identified that the 22-mer corresponded to 5-nt being trimmed from the 5' end of the guide strand rather than 6-nt as expected. They surmised that there may be link between Dicer processing and generation of a guide strand more favourable for strand selection and gene silencing.

For standard 5' RACE, which uses Sanger sequencing to confirm the sequence of 5' RACE clones, data is typically reported as expected cleavages per total clones with 5' end at the expected cleavage site (Davis et al., 2010; Clark et al., 2013; Sakurai et al., 2013). Typically, 5' end counts outside of the expected observed cleavage site is not reported, so that, depending on the number of clones sequenced, a true shift in RISC cleavage site may not be detected.

Interestingly, a group at Dicerna Pharmaceuticals, Inc., presenting their work evaluating a lipid-nanoparticle (LNP)-formulated 2'-O-Me modified DsiRNA in vivo confirmed a RNAi mechanism of action using RACE-Seq. Despite showing the expected canonical processing of their DsiRNA by Dicer to generate a 21-mer in the introduction of the poster, the RACE-Seq data (a peak reporting >90% cleavage) prompted the acceptance that a 22-mer was processed by Dicer with a shift in the expected cleavage position and added the statement; "The location of Ago2 cleavage predicts the sequence of the mature siRNA (post-Dicer processing)". This statement, although obscurely buried on a poster, encompasses the greatest opportunity for RACE-Seq to enter the RNAi field as a standard assay for confirming a RNAi mechanism of action. The poster was accessed on 30 May 2017 (but has since been removed from the internet). Some of the information is available in a presentation from the 12th Annual meeting of the Oligonucleotide Therapeutics Society, 28 September 2017 (http://files.shareholder.com/downloads/AMDA-2IH3D0/5213734735x0x909887/15FC393C-6764-4983-87C0-64DED88FD75D/DCR-MYC _Presentation_for_OTS _2016 _FINAL.pdf).

This chapter represents the first observation of RISC-induced canonical cleavage of target by an non-cleaved 27-mer Dicer substrate guide strand. The current assumption is that the entire 27-mer was loaded to RISC which is supported by the work of Snead et al*.,* (2013) who undertook to investigate the basis for improved gene silencing by Dicer substrate siRNA. They identified that the 27-mer bottom strand of a Dicer substrate was purified from immuno-precipitated AGO2. Much of the early work investigating 27-mer siRNA focused on determining the optimal structural elements for enhanced potency and used electrospray ionization mass spectrometry (ESI-MS) to examine the oligonucleotides as well as *in vitro* cleaved Dicer products (Kim et al., 2005; Rose et al., 2005; Amarzguioui et al., 2006). Further support for the theory of unprocessed 27-mer guide strand comes from the work by Salomon et al., (2010) who showed that an unmodified 25-mer blunt end duplex cleaved its target at the expected position between bases 10-11 from the end of the 25-mer guide strand. They employed an *in vitro* mRNA cleavage assay incubating ^{32}P-labelled target RNA and immune-precipitated AGO2 together and then analysing the cleaved products for the expected size fraction (Salomon et al., 2010). In conclusion, for this experiment, RACE-Seq predicts a RISC-induced on-target mechanism of action for two different guide strand species of DsiRNA-TTR.

Currently, validation of RNAi drug activity is driven by reporting of clinically relevant information such as efficacy, reduction in dosage, increased bioavailability, improved targeting, reduced immune activation, no off-target effects and safety. The standard 5' RACE assays dictated reporting an expected RISC cleaved product based on the design of the RNAi bioactive. In contrast, RACE-Seq provides the opportunity to characterise RISC activity in an unbiased manner. For Dicer substrate RNAi triggers, RACE-Seq allows a means of reverse predicting Dicer behaviour, as RACE-Seq defines the mature siRNA and thus predicts Dicer cleavage action.

6 RACE-SEQ REPORTS THE ACTIVE SIRNA SEQUENCE

6.1 Introduction

Our understanding of AGO2-mediated RNAi activity is that RISC cleaves target RNA opposite bases 10-11 of the siRNA guide strand, when counting from the 5' end of the hybridized strand (Gu et al., 2012; Liu et al., 2015). This implies that the sequence of the cleavage product can predict the sequence (or at least the 5' end) of the active siRNA guide strand. Currently, RML-RACE with Sanger sequencing remains the gold standard for confirming a direct mechanism of action of AGO2-mediate RISC cleavage. However, the much lower read throughput of Sanger sequencing may result in a misleading or narrowed interpretation of RISC activity, as currently the method is used to confirm a pre-conceived expectation for the behaviour of a designed RNAi-based trigger (Frank-Kamenetsky et al., 2008; Judge et al., 2009; Davis et al., 2010). While AGO2-RISC cleavage remains precise, 'imprecise' Dicer cleavage is well documented (Saayman et al., 2010; Starega-Roslan et al., 2011; Harwig et al., 2015) and a number of attempts have been made to define the precise sequence and/or structural characteristics of Dicer substrate including shRNAs (Rose et al., 2005; Dallas et al., 2012). Results from Chapter 5, regarding the expected vs RACE-Seq predicted cleavage site for DsiRNA-TTR, prompted the notion that RACE-Seq could identify the sequence of the RNAi trigger.

In the following chapter, RACE-Seq was applied in an unbiased manner to investigate the potential RISC-induced cleavage products of six RNAi bioactive oligonucleotides targeting the HCV replicon genome. This information was then used to predict the sequence of the active siRNA trigger and to hypothesize Dicer processing activity.

6.2 Results

6.2.1 Data output per sample and read depth

In Chapter 4, two independently prepared RACE-Seq sequencing runs which were conducted at separate time points is described. The total number of reads for each sample in the runs varied greatly across the 14 samples (Table 6.1 and 6.2). All RACE-Seq analysis was carried out by filtering reads to have adapter+20 nt as indicated in Figure 5.5. The data filtering protocol included searching for the adapter in both directions, reverse complementing the reads with the adapter sequence in the reverse orientation, pooling the reads from the two files, filtering reads for the target sequence by trimming expected sequences from the 3' end of reads before trimming reads to length adapter+20-nt and then this final short read file was used as input to the RACE-SEQ-lite pipeline (Theotokis et al., 2017) which removes the RNA adapter sequence and maps the reads to the reference to determine the 5' end counts.

The peak information is most easily interpreted as a bar graph of 5' end counts that align at the various positions within the expected RISC hybridization site. However, variation in read amount can make it difficult to assess each sample on its individual merit. Presenting the data as a percent of aligned 5' ends solves this by allowing each sample to be assessed individually, but uses a common factor, percentage aligned reads. For example, siRNA22(2) had 71,152 reads at the expected cleavage site, while sample siRNA22(1) had almost twice as many reads (135,571) at the expected cleavage site (Figure 6.1 A and B). Comparing them side by side may misinterpret that the sample with fewer reads is a 'poorer' sample. However, when each of the datasets is presented as a percentage of aligned reads, siRNA22(1) had 43.8% of 5' ends at the expected cleavage site and siRNA22(2) had 50.8% of all aligned 5' ends at the expected cleavage site.

When presented as the proportion of aligned reads, read amount becomes less of an obstacle to data interpretation. By the same token, only presenting the data as a percentage of aligned reads, without disclosure of total filtered reads (ie. the number of reads that were input for alignment), can mask data which is truly less robust. For example, RACE-Seq analysis for samples shRNA6(1) and shRNA6(2) reported 36.2% and 40.1% of 5' ends aligned at the expected cleavage position for the respective samples. If presented only as percentage of 5' ends aligned to the target, without disclosure of the total reads (5,812 and 29,934 reads respectively) and total number of reads after cleaning the data (1,988 and 19,409 reads), the fact that both samples had lower than expected number of reads for each of the respective sequencing runs (Table 6.2) would not be disclosed. Ultimately, the total number of reads obtained for each sample in the HCV knockdown RACE-Seq assays had no impact on confirming the expected RISC cleavage position. However, this hinges off good quality control checkpoints during sample preparation (ie. primer validation for 5'

RML-RACE, verification of 5’ RACE-PCR product by agarose gel analysis, size selection to eliminate unwanted fragments from NGS library preparation and sensitive validation of NGS sample fragment size by electronic fragment analysis.

Thus, regardless of the variability in the number of reads obtained for each sample, RACE-Seq robustly reported the RISC-induced cleavage site (Figure 6.1 and Figure 6.3). Additionally, the independent experiments reported very similar peak patterns for the same RNAi-based trigger. This indicates that the various downstream protocols for this RACE-Seq assay, from adapter ligation through to size selection of NGS samples, was robust. For shRNA6, despite sub-optimal knockdown conditions, ~60% knockdown when using 0.5 nmol/l effective concentration (Table 4.1), RACE-Seq still reported the expected cleavage site (Table 6.4, Figure 6.3 C). Interestingly, when the concentration of shRNA6 analogue was increased from 0.5 nmol/l to 1.0nmol/l (Table 4.1), an increase in silencing resulted in a greater number of RACE-Seq reads being obtained (Table 6.2, Figure 6.3 C and D). This is as expected as the increased dose would have resulted in improved knockdown and an increase in RISC cleaved products which could be captured by RML RACE and reported by NGS sequencing. Thus, well defined RML RACE assays and quality controlled NGS library preparation generated a robust RACE-Seq assay.

Table 6.1 Read statistics for siRNA6, siRNA19 and siRNA22.

							Reads that align to target	
	Total Reads	Adapter+20nt	Reads with adapter	<10 bp	Total aligned reads	Unaligned reads	% of total reads	% of reads with adapter
siRNA22 (1)	613,557	441,731	392,102	1	309,725	82,376	50.5	79.0
siRNA22 (2)	478,979	271,934	206,110	3	143,986	62,121	30.1	69.9
siRNA6 (1)	216,114	179,132	107,903	3,035	95,286	9,582	44.1	88.3
siRNA6 (2)	151,825	120,459	107,668	113	100,028	7,527	65.9	92.9
siRNA19 (1)	624,683	490,279	422,419	3,226	374,938	44,255	60.0	88.8
siRNA19 (2)	153,319	130,606	115,298	408	106,143	8,747	69.2	92.1
siRNA19 (3)	443,420	329,909	300,926	5,227	259,579	36,120	58.5	86.3

Table 6.2 Read statistics for shRNA6, shRNA19 and shRNA22.

	Total Reads	Adapter+20nt	Reads with adapter	<10 bp	Total aligned reads	Unaligned reads	Reads that align to target	
							% of total reads	% of reads with adapter
shRNA22 (1)	359,022	258,543	240,931	0	171,196	69,735	47.7	71.1
shRNA22 (2)	468,498	223,793	187,621	5	139,912	47,704	29.9	74.6
shRNA6 (1)	5,812	3,750	1,988	142	1,632	214	28.1	82.1
shRNA6 (2)	29,934	21,775	19,409	117	17,280	2,012	57.7	89.0
shRNA19 (1)	615,160	437,169	344,639	1,171	294,357	49,111	47.9	85.4
shRNA19 (2)	162,539	144,082	131,871	247	121,055	10,569	74.5	91.8
shRNA19 (3)	701,310	509,694	450,578	2,722	383,353	64,503	54.7	85.1

6.2.2 RACE-Seq confirms the expected cleavage site for siRNA22

Previous RACE-Seq analysis for siRNA22 using Illumina high throughput sequencing reported the expected RISC cleaved site at position T=9489 on the sense strand of the HCV replicon RNA genome (Denise et al., 2014). Using the legacy RACE-Seq assay, the expected cleavage site for siRNA22 was confirmed when sequencing on the Ion Torrent PGM (Chapter 3). Additionally, a second peak 3-nt upstream of this expected cleavage site was reported at 17% of aligned reads at position A=9486 (Chapter 3, Figure 3.13). No correlation between the Illumina data from Denise et al. and the Ion Torrent legacy RACE-Seq siRNA22 data could be determined. In the absence of corroborating evidence that this second peak was potentially generated by RISC-induced activity, this second site was dismissed as undetermined origin.

For the long amplicon RACE-Seq assays, the FASTQ file for each sample was pre-filtered as outlined previously and then aligned to the HCV replicon reference genome using the RACE-SEQ-LITE pipeline. For siRNA22, the RISC hybridization site lies at position 9479 to 9498 on the replicon genome. For each of the samples, despite the difference in total read count (Table 6.1), each of the independently prepared and sequenced siRNA22 samples reported the expected peak at T = 9489 (Figure 6.1). When the peak profiles for each sample was analysed as a percent of aligned reads, sample siRNA22(1) reported 43.8% of reads at the expected cleavage site and sample siRNA22(2) reported 50.8% of reads at the expected cleavage site (Table 6.3). In each case, the expected peak was reliably detected by RACE-Seq and confirmed the sequence of the guide strand and that the siRNA22 molecule had performed in the expected manner for the design of the duplex (Figure 6.1 A and B, Figure 6.2 A).

However, a second 5' end peak at 2-nt upstream of the expected cleavage site was also reported for both samples (Figure 6.1 A and B), which had 14.5% of reads for siRNA22(1) and 6.8% of reads for siRNA22(2) (Table 6.3). The RACE-Seq legacy assay for siRNA22 had reported a second peak at 3-nt (A=9476) upstream of the expected RISC cleavage site, but in the absence of corroborating evidence, was dismissed as being unlikely to have been generated by RISC activity. However, since this A=9487 peak was reported by both samples as the second highest count, this warranted a closer look at the peak profiles and speculation on their possible origin. If the secondary peak is to be taken as real reporting of RISC activity, this would infer that an active siRNA22 species of compatible sequence and possessing the necessary 2-nt 3' overhang structure for AGO2 recognition had been generated *in vivo* or a part of the siRNA22 sample. Dicer is not known to process 21-mer siRNA such as siRNA22, and only a small set of published work that included a 21-mer in a Dicer cleavage assay could be found. For example, an *in vitro* Dicer assay was performed by incubating a 21-mer, 23-mer, 25-mer, and 27-mer siRNA with recombinant Dicer for 24 hours. The products were separated on a polyacrylamide gel, stained and compared to a set of untreated samples. The 21-mer remained unchanged, but all the other duplex sizes showed a shift in size that corresponded to the 21-mer untreated sample position in the gel (Kim et al., 2005). Whether this second upstream peak is induced by a siRNA22 species remains to be further investigated.

Both the siRNA22 RACE-Seq samples also reported a peak at A=9479, G=9481 and T=9483 (Figure 6.1 A and B). The nature by which these peaks were derived can only be speculated to be due to some other biological processing/cleavage of RNA. They are unlikely to be as a result of a RISC-induced siRNA22 species as the duplex length required would be too short for AGO2-RISC recognition.

6.2.3 RACE-Seq confirms the expected cleavage site for shRNA22

Both the shRNA22(1) and shRNA22(2) RACE-Seq samples reported the expected cleavage position, with 53.7% and 51.2% of filtered reads aligning to position T=9489 for respective samples (Table 6.3). The shRNA22 synthetic hairpin molecule was designed to mimic the shRNA22 pro-drug expressed from TT-034. The hairpin is designed as a reverse orientated shRNA, meaning that the guide strand is on the 5' side of the hairpin structure (Lavender et al., 2012). Therefore, any unexpected cleavage events by Dicer processing would not alter the start position of the guide strand as the 5'end of the guide strand remains unaffected (Figure 6.4 A). The expectation is that the first nucleotide on the 5' end of the shRNA is the first nucleotide of the active guide strand. A count of 10 from the end of the guide strand then positions AGO2-RISC cleavage at the expected position for cleavage of the target RNA, giving a peak of 5' ends after RACE-Seq (Figure 6.3 A and B).

For shRNA22(2) a second peak, is reported as 6.8% of aligned reads (Table 6.3, A=9487) at 2-nt positions upstream of the expected cleavage site. However, some caution must be taken in interpreting this second peak as a RISC-driven cleavage observation. One possible explanation for the second peak could be cross contamination of PCR samples, as the siRNA RACE-Seq samples do show a second peak at this position. The siRNA22(2) and shRNA22(2) samples were being processed at the same time and even a very small amount of cross contamination would be amplified and detected by NGS. Since this second peak was only observed in one of the samples, and was not seen in shRNA22(1), this gives less credence to this peak deriving from RISC-associate activity. Further, it is unlikely that a suitable guide strand could be generated from shRNA22 due to the length of the shRNA and the orientation of the guide strand.

6.2.4 RACE-Seq confirms the expected cleavage site for siRNA6 and identifies the trigger guide strand sequence after Dicer cleavage of shRNA6

Both of the RACE-Seq samples for siRNA6 confirmed the expected RISC cleavage position as expected from the design of the synthetic siRNA6 molecule (Figure 6.1 C and D, Figure 6.2 B). Cleavage of the HCV replicon genome occurred at position T=282 with 75.1% of 5' ends obtained for siRNA6(1) and 73.4% of 5' ends for siRNA6(2) (Table 6.4).

The sequence for the siRNA6 molecule was chosen based on the most common siRNA sequence observed when the small RNAs were sequenced after transfection of TT-034 to Huh7/con1b cells (Denise et al., 2014). However, the guide strand sequence of the synthetic siRNA6 molecule has at its 5' end 1-nt from the shRNA6 hairpin loop structure and the 3' end of the passenger stand has 3-nt of the loop structure (Figure 6.2 B), and the siRNA6 molecule is therefore not the same as the expected siRNA that would be generated by Dicer cleavage of shRNA6 (Figure 6.4 B).

Since the shRNA6 molecule has the typical 21-bp stem-loop structure, the 'loop counting rule' could be applied when predicting Dicer cleavage activity (Gu et al., 2012). The 'loop counting rule' states that for a duplex shRNA of 21-bp, Dicer recognises the single stranded region of the loop structure and cleaves the guide strand 2-nt from the start of the loop (Figure 6.4 B). RACE-Seq results for shRNA6 confirm the expected cleavage position (Figure 6.3 C and D) and obeyed the 'loop counting rule'. Deriving further insights into shRNA6 analogue processing is restricted due to much lower number of total reads obtained for these two samples. Despite this low number of reads, particularly for shRNA6(1) which had only 5,182 reads in total (Table 6.2) and 590 reads at the expected cleavage position (Figure 6.3 C), RACE-Seq confirmed the expected cleavage position. Thus, well-designed and validated RACE-Seq assays have the potential to be utilized for targets with low expression.

Table 6.3 Percent aligned reads for siRNA22 and shRNA22 RACE-Seq assays

	Hybridized guide strand position	Nucleotide	siRNA22(1)	siRNA22(2)	shRNA22(1)	shRNA22(2)
3'	9478	C	0.0	0.2	0.0	0.2
	9479	A	4.1	0.8	0.0	0.8
	9480	A	0.3	0.1	0.0	0.1
	9481	G	5.5	1.8	0.0	1.8
	9482	C	0.0	0.2	0.0	0.2
	9483	T	4.2	1.4	0.0	1.4
	9484	C	0.0	0.0	0.0	0.0
	9485	A	0.0	0.0	0.0	0.0
	9486	A	0.4	0.2	0.0	0.2
	9487	A	14.5	6.8	0.1	6.8
	9488	C	0.7	1.8	1.9	1.8
	9489	T	43.8	50.8	53.7	51.2
	9490	C	14.2	25.7	37.3	25.3
	9491	A	7.9	5.9	3.2	5.9
	9492	C	0.7	1.6	1.8	1.6
	9493	T	0.0	0.2	0.1	0.2
	9494	C	3.6	1.2	1.4	1.2
	9495	C	0.0	0.1	0.0	0.0
	9496	A	0.0	0.0	0.0	0.0
	9497	A	0.0	0.1	0.4	0.1
5'	9498	T	0.0	0.1	0.0	0.1

RACE-Seq datasets were pre-filtered to specifically select for the target sequence and then trimmed to adapter+20-nt length. These short reads were input to the RACE-SEQ-lite analysis pipeline which performed the alignment reported the 5' end counts as a percent of total aligned reads. Both siRNA22 and shRNA22 reported the highest percent of 5' ends at position 9489 = T (in red), corresponding to a count of 10 from the 5' end of the siRNA or expected dicer processed shRNA22 guide strand.

Table 6.4 Percent aligned reads for siRNA6 and shRNA6 RACE-Seq assays

	Hybridized guide strand position	Nucleotide	siRNA6(1)	siRNA6(2)	Hybridized guide strand position	Nucleotide	shRNA6(1)	shRNA6(2)
3′	272	C	0.0	0.9	269	T	3.5	4.8
	273	G	0.0	0.2	270	C	2.5	3.4
	274	A	0.4	0.4	271	G	0.2	0.7
	275	A	1.6	0.7	272	C	1.1	1.4
	276	A	0.0	0.6	273	G	0.0	0.6
	277	G	0.0	0.2	274	A	0.3	0.2
	278	G	0.0	0.2	275	A	0.2	0.1
	279	C	0.8	1.2	276	A	0.6	0.3
	280	C	1.5	0.9	277	G	0.0	0.4
	281	T	3.3	1.8	278	G	0.0	1.0
	282	T	75.1	73.4	279	C	36.2	40.1
	283	G	0.9	0.6	280	C	9.1	6.0
	284	T	0.0	1.0	281	T	3.1	1.3
	285	G	0.0	0.6	282	T	1.9	1.3
	286	G	0.0	0.6	283	G	0.0	0.1
	287	T	0.0	0.7	284	T	0.0	0.7
	288	A	0.0	0.4	285	G	0.0	0.4
	289	C	2.7	1.0	286	G	0.0	1.1
	290	T	0.0	0.1	287	T	0.4	1.5
5′	291	G	0.0	0.2	288	A	0.0	1.0

For the RACE-Seq datasets of siRNA6 and shRNA6 that were trimmed to adapter+20-nt length and aligned to the HCV reference using the RACE-SEQ-lite analysis pipeline, two different cut sites on the target was indicated by 5’ end counts. While siRNA6 (left and) reported the highest number of aligned reads at 282 = T (red), the shRNA analogue reported the highest amount of reads at position 279 = C (red). In each case, this corresponded to canonical RISC-induced cleavage by the expected guide sequence.

Table 6.5 Percent aligned reads for siRNA19 and shRNA19 RACE-Seq assays

	Hybridized guide strand position	Nucleotide	siRNA19(1)	siRNA19(2)	siRNA19(3)	shRNA19(1)	shRNA19(2*)	shRNA19(2)
3'	9095	G	0.0	0.0	0.0	0.0	0.0	0.1
	9096	T	0.0	0.0	0.0	0.1	0.0	0.2
	9097	C	0.2	0.0	0.0	0.6	0.1	0.3
	9098	A	0.0	0.0	0.0	0.2	0.0	0.1
	9099	A	0.0	0.0	0.0	0.0	0.0	0.0
	9100	T	0.0	0.0	0.0	0.0	0.0	0.1
	9101	T	0.0	0.0	0.0	0.0	0.0	0.1
	9102	C	0.0	0.0	0.0	0.0	0.0	0.1
	9103	C	0.0	0.1	0.0	0.5	0.1	0.2
	9104	T	14.5	17.3	10.9	17.4	19.4	12.0
	9105	G	15.7	37.8	15.0	35.4	55.8	28.9
	9106	G	1.8	3.0	1.7	1.0	1.7	0.9
	9107	C	18.8	12.2	11.6	6.8	10.4	3.8
	9108	T	18.8	15.0	10.6	0.1	1.2	1.4
	9109	A	11.6	6.4	3.9	3.1	1.1	1.2
	9110	G	5.4	2.2	5.2	2.4	0.8	0.9
	9111	G	0.9	0.5	1.6	5.7	0.4	0.4
	9112	C	1.5	0.1	1.6	2.3	0.2	0.3
	9113	A	0.3	0.2	0.0	1.1	0.3	0.2
	9114	A	0.2	0.1	0.0	1.8	0.3	0.2
	9115	C	0.8	0.3	0.0	2.2	0.6	0.4
	9116	A	0.0	0.1	0.0	0.7	0.1	0.2
5'	9117	T	0.0	0.1	0.0	0.0	0.3	0.2

The RACE-Seq datasets of siRNA19 and shRNA619 were trimmed to adapter+20-nt length and aligned to the HCV reference using the RACE-SEQ-lite analysis pipeline. The shRNA19 analogue reported four positions with high amount of aligned 5' ends. The fist corresponded to a count of 10 from the 5' end of the hybridized guide strand at position 9108 = T (red). Other 5' end aligned counts corresponded to 9107 = C, 9155 = G and 9107 = T (red). For a canonical RISC cleavage mechanism to be maintained, this would necessitate that the siRNA19 analogue was processed *in vivo* to the appropriate 5' start site. The shRNA19 analogue generated two positions of 5' aligned reads at 9105 = G and 9104 = T (red).

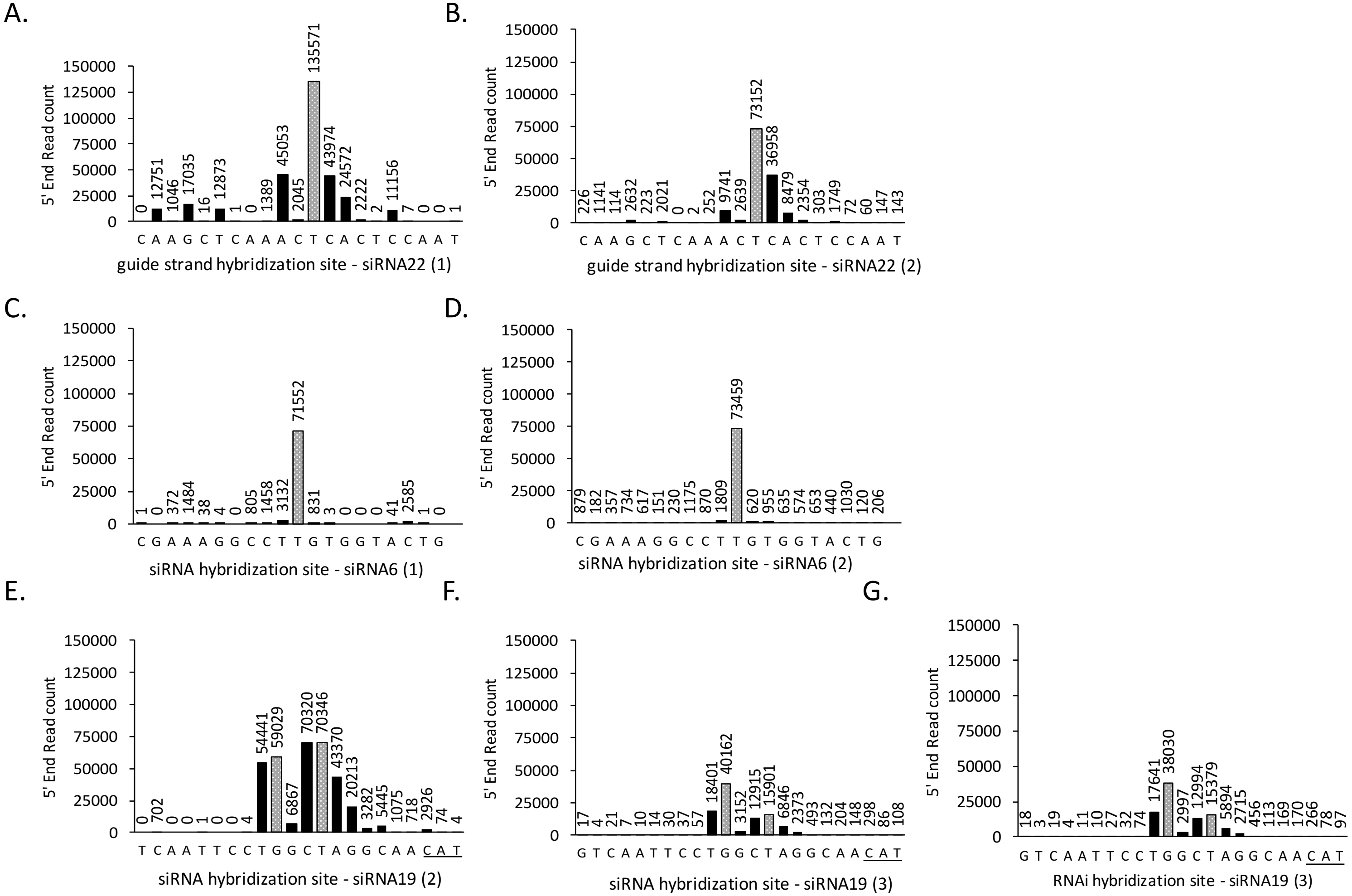

A.
5' End Read count
150000
125000
100000
75000
50000
25000
0
0 12751 1046 17035 16 12873 1 0 1389 45053 2045 135571 43974 24572 2222 2 11156 7 0 0 1
C A A G C T C A A A C T C A C T C C A A T
guide strand hybridization site - siRNA22 (1)
B.
5' End Read count
226 1141 114 2632 223 2021 0 2 252 9741 2639 73152 36958 8479 2354 303 1749 72 60 147 143
C A A G C T C A A A C T C A C T C C A A T
guide strand hybridization site - siRNA22 (2)
C.
5' End Read count
1 0 372 1484 38 4 0 805 1458 3132 71552 831 3 0 0 0 41 2585 1 0
C G A A A G G C C T T G T G G T A C T G
siRNA hybridization site - siRNA6 (1)
D.
5' End Read count
879 182 357 734 617 151 230 1175 870 1809 73459 620 955 635 574 653 440 1030 120 206
C G A A A G G C C T T G T G G T A C T G
siRNA hybridization site - siRNA6 (2)
E.
5' End Read count
0 702 0 0 1 0 0 4 54441 59029 6867 70320 70346 43370 20213 3282 5445 1075 718 2926 74 4
T C A A T T C C T G G C T A G G C A A C A T
siRNA hybridization site - siRNA19 (2)
F.
5' End Read count
17 4 21 7 10 14 30 37 57 18401 40162 3152 12915 15901 6846 2373 493 132 204 148 298 86 108
G T C A A T T C C T G G C T A G G C A A C A T
siRNA hybridization site - siRNA19 (3)
G.
5' End Read count
18 3 19 4 11 10 27 32 74 17641 38030 2997 12994 15379 5894 2715 456 113 169 170 266 78 97
G T C A A T T C C T G G C T A G G C A A C A T
RNAi hybridization site - siRNA19 (3)

Figure 6.1 RACE-Seq detection of 5' end peaks for siRNA-mediated cleavage of HCV replicon RNA presented as percentage of total aligned reads.

For two independent RACE-Seq assays, the expected cleavage site –light coloured bar- was identified with the highest number of 5’ ends for the siRNA22 samples (A) and (B). For siRNA6 (C and D), RACE-Seq reported the expected cleavage site based on canonical RISC cleavage. For siRNA19, RACE-Seq for three independent samples reported a peak pattern consisting of two pairs of peaks (E, F and G). The light-coloured peaks are the expected cleavage positions based on an unprocessed siRNA19 and classic ‘loop’ counting rule Dicer processed siRNA guide strand-mediate cleavage. The black peaks at +1 of the light-coloured peaks are postulated to be novel 5’ ends generated from additional siRNA19 Dicer processed species.

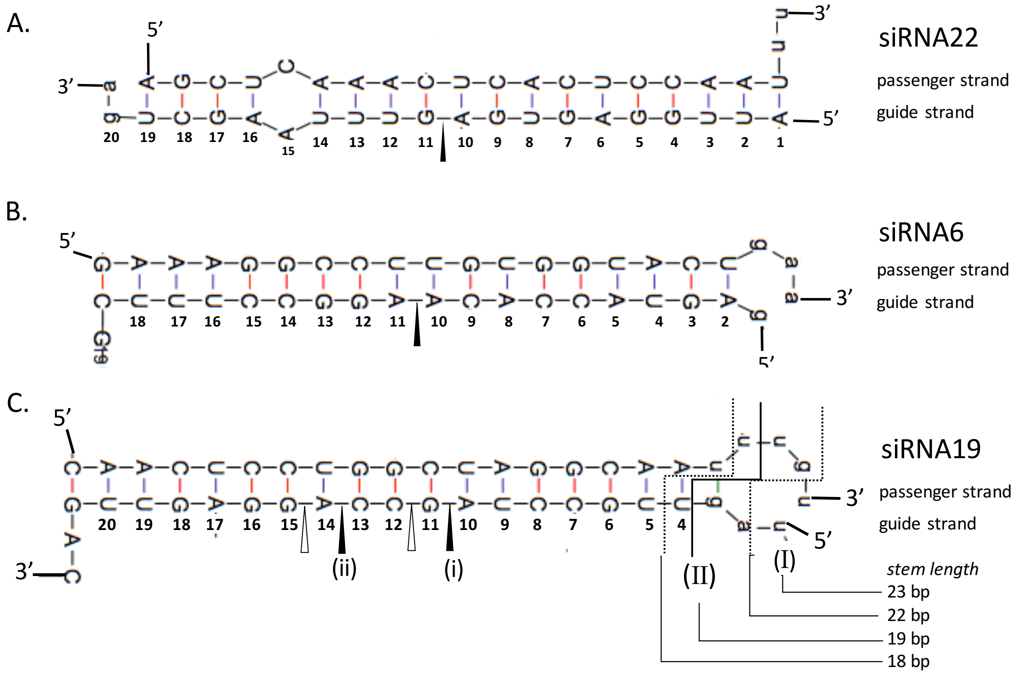


Figure 6.2 Illustration of duplex siRNA Illustration of predicted and RACE-Seq observed cleavage positions for three siRNAs targeting the sense strand of HCV replicon genome.

(A) illustrates the hybridization of the guide and passenger RNA strands for siRNA22, siRNA6 (B) and siRNA19 (C). In each case, the black triangle indicated the expected AGO2 cleavage position. For siRNA19, the black triangles represent the proposed AGO2 cleavage sites based on the results from RACE-Seq data analysis and correspond to positions (I) and (II) as the start of the siRNA guide strands. The open triangles represent additional proposed AGO2 cleavage sites based on RACE-Seq data analysis with proposed corresponding start sites for guide strands at the dotted lines. The nucleotides are counted from the 5’ end of the guide strand or proposed end of cleaved strand. Lowercase letters = sequences that alogn to the loop the shRNA partner.

6.2.5 RACE-Seq reports two active siRNAs processed from shRNA19

The sequence for shRNA19 is the same as the shRNA sequence that is expected to be expressed from TT-034. In three separate experiments, RACE-Seq reported two distinct 5' end peaks around the expected cleavage site (Figure 6.3 E, F and G). The primary peak was found at position G=9105, with 35.4%, 55.8% and 28.9% of 5' ends aligning to this expected cleavage site for each of the respective samples (Table 6.5). For all samples, a second peak at T=9104 with 17.4%, 19.4% and 12.0% of aligned reads for the respective samples was also identified (Table 6.5). This confirms the previous result for 5' RACE-PCR and Sanger sequencing when the cleavage products of TT-034 were assessed (Denise et al., 2014). The same two 5' ends were identified from cleavage products.

Since the second peak is positioned upstream of the primary peak, RACE-Seq data analysis led to the idea that at least two, highly active siRNAs must be processed from shRNA19 by Dicer cleavage. The sequence for shRNA19 was put into the UNAFold Web Server (http://unafold.rna.albany.edu/) to generate the predicted folded hairpin structure for shRNA19 and revealed an interesting characteristic regarding the hairpin loop sequence. The loop sequence of shRNA19 analogue is (uuuguguag) and is different from the shRNA22 and shRNA6 loop sequence which is (gaagcuug) (Figure 6.4). The predicted structure for shRNA19 indicated that the loop structure could be collapsed (McIntyre and Fanning, 2006), such that instead of the expected 9-nt shRNA loop based on the design of the molecule, a 5-nt loop occurs (Figure 6.4 B). Since the second cleavage product is extended by just 1 nucleotide, this must have been generated from a shRNA19 molecule with 7 nucleotides (Figure 6.4 C).

The 'loop counting' rule can be used to predict the siRNA guide strand sequences and the expected cleavage products that would be generated from Dicer cleavage of shRNA19 when the two possible loop configurations of shRNA19 are considered. The collapsed loop would generate the primary cleavage product (highest number of read), while the second structure accounts for the second peak (Figure 6.3 E, F, G and Figure 6.4 C). It is thus likely that shRNA19 exists as a mix of two species in the biological habitat, with some molecules having a 5-nt loop and others with a 7-nt loop.

Four primary conclusions may be drawn from the RACE-Seq analysis of shRNA19; (1) that RACE-Seq reliably detects the expected AGO2-RISC cleaved site, (2) that RACE-Seq detects a second cleavage product, (3) RACE-Seq drives the prediction of the sequence of a second active siRNA19 species that was generated by Dicer cleavage of shRN19 and (4) analysis of the loop structure supported the notion of a dual structure for shRNA19.

6.2.6 RACE-Seq reports multiple active siRNA species from siRNA19

In three independent RACE-Seq assays to assess AGO2-mediated cleavage of HCV replicon genome triggered by siRNA19, two distinct clusters of peaks, consisting of two peaks each was observed. The peaks occurred at nucleotide positions T = 9104/G = 9105 and C = 9107/T = 9108 (Figure 6.1 E, F and G). Although the rank order of the 5' end peaks varied for the three samples, these four peaks consistently ranked as the top percentage of aligned 5' ends for each of the samples (Table 6.5). This observation of a pattern across all three siRNA19 samples led to closer examining of the siRNA19 sequence.

The siRNA19 sequence was compared to the shRNA19 sequence and the sequence of its target site. In doing so, siRNA19 was identified as a 23-mer siRNA that differed in sequence from the siRNAs that are expected to be processed from shRNA19 by Dicer cleavage. The sequence for siRNA19 was chosen as the most prominent siRNA19 sequence after TT-034 transfection to Huh7/con1b cells and sequencing of the small RNAs (Denise et al., 2014). Additionally, the siRNA19 molecule has 3 nucleotides from the shRNA19 hairpin loop on the 5' end of the guide strand (uuugugu<u>ag</u>, underlined) and opposite this end, the passenger strand also has 5 nucleotides of the loop structure sequence (**uuugu**guag, in bold) at the 3' end of the passenger strand (Figure 6.4 C).

Since siRNA19 is a 23-mer siRNA, and has one end of undefined structure, this siRNA is a possible substrate for Dicer (Kim et al., 2005) However, care should be taken not to over-interpret what looks like a hairpin structure at the 5' end of the guide strand as this may be an artefact of the UNAFold software. In order to view the potential active siRNA19 sequences that generated the observed RACE-Seq peaks, the peak information was overlaid onto the siRNA19 guide strand. The RACE-Seq peak profiles indicate that there are four possible active siRNA guide sequences, which correspond to 23-mer, 22-mer, 19-mer and 18-mer siRNAs (Figure 6.2 C). It remains to be determined whether active siRNAs are being generated from this siRNA19 molecule by either Dicer recognition and cleavage or not cleavage or some other cellular cleavage activity.

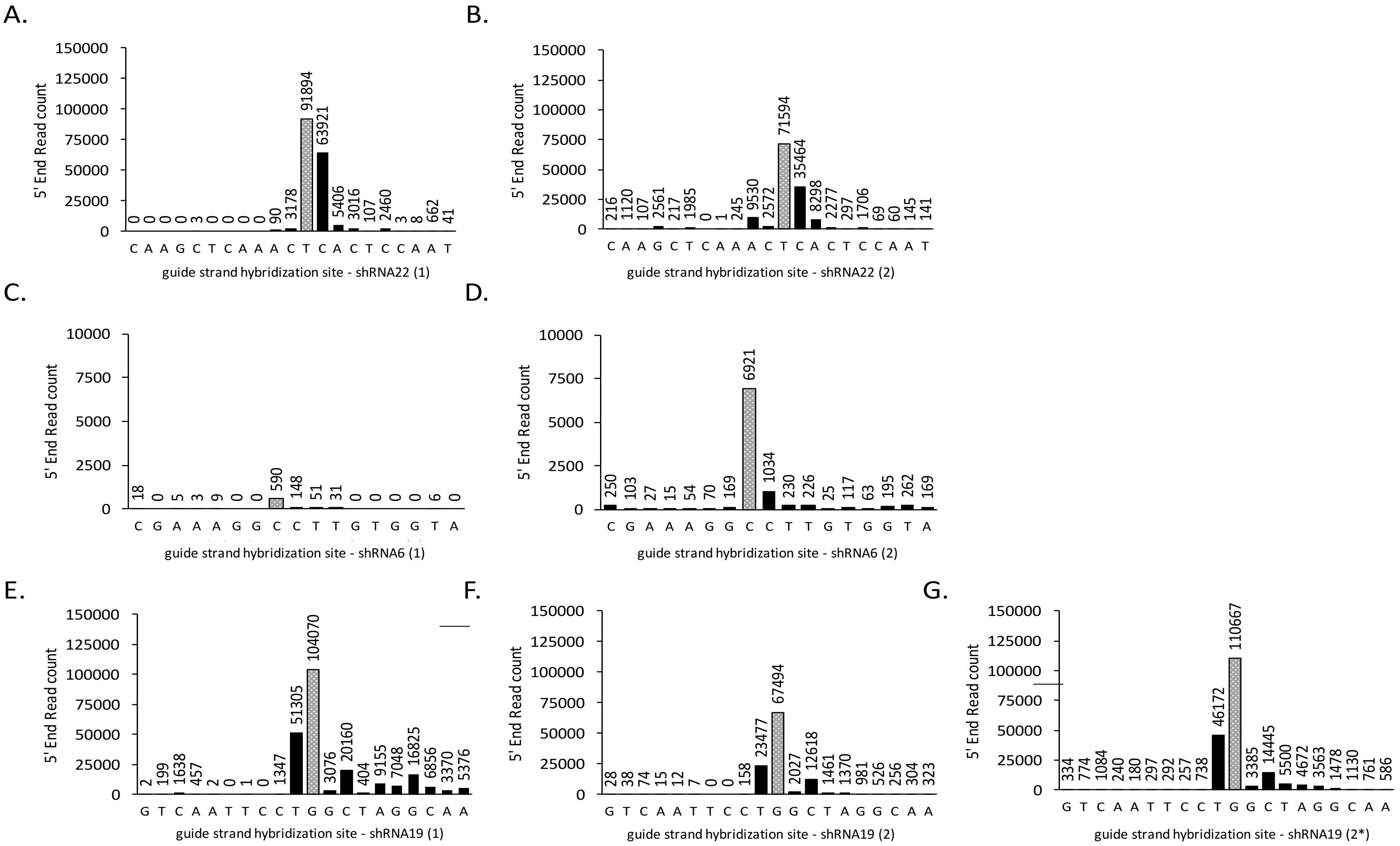

A.
150000
125000
100000
75000
50000
25000
0
5' End Read count
0 0 0 0 3 0 0 0 0 90 3178 91894 63921 5406 3016 107 2460 3 8 662 41
C A A G C T C A A A C T C A C T C C A A T
guide strand hybridization site - shRNA22 (1)
B.
150000
125000
100000
75000
50000
25000
0
5' End Read count
216 1120 107 2561 217 1985 0 1 245 9530 2572 71594 35464 8298 2277 297 1706 69 60 145 141
C A A G C T C A A A C T C A C T C C A A T
guide strand hybridization site - shRNA22 (2)
C.
10000
7500
5000
2500
0
5' End Read count
18 0 5 3 9 0 0 590 148 51 31 0 0 0 0 6 0
C G A A A G G C C T T G T G G T A
guide strand hybridization site - shRNA6 (1)
D.
10000
7500
5000
2500
0
5' End Read count
250 103 27 15 54 70 169 6921 1034 230 226 25 117 63 195 262 169
C G A A A G G C C T T G T G G T A
guide strand hybridization site - shRNA6 (2)
E.
150000
125000
100000
75000
50000
25000
0
5' End Read count
2 199 1638 457 2 0 1 0 1347 51305 104070 3076 20160 404 9155 7048 16825 6856 3370 5376
G T C A A T T C C T G G C T A G G C A A
guide strand hybridization site - shRNA19 (1)
F.
150000
125000
100000
75000
50000
25000
0
5' End Read count
28 38 74 15 12 7 0 0 158 23477 67494 2027 12618 1461 1370 981 526 256 304 323
G T C A A T T C C T G G C T A G G C A A
guide strand hybridization site - shRNA19 (2)
G.
150000
125000
100000
75000
50000
25000
0
5' End Read count
334 774 1084 240 180 297 292 257 738 46172 110667 3385 14445 5500 4672 3563 1478 1130 761 586
G T C A A T T C C T G G C T A G G C A A
guide strand hybridization site - shRNA19 (2*)

Figure 6.3 RACE-Seq detection of 5' end peaks for siRNA-mediated cleavage of HCV replicon RNA presented as percentage of total aligned reads.

For two independent RACE-Seq assays, the expected cleavage site –light coloured bars- was identified with the highest number of 5’ ends for the shRNA22 samples (A) and (B). For siRNA6 (C and D), RACE-Seq reported the expected cleavage site based on the expected Dicer processed guide strand sequence. For shRNA19, RACE-Seq for three independent samples reported a peak pattern consisting of two distinct peaks (E, F and G). The light-coloured peaks are the expected cleavage position based on the ‘loop’ counting rule of Dicer processed shRNA19 with a 5-nucleotide loop structure, while the black bar at +1(T) of the expected cleavage site corresponds to a guide strand that was Dicer processed from a shRNA19 with a 7-nucleotide loop structure.

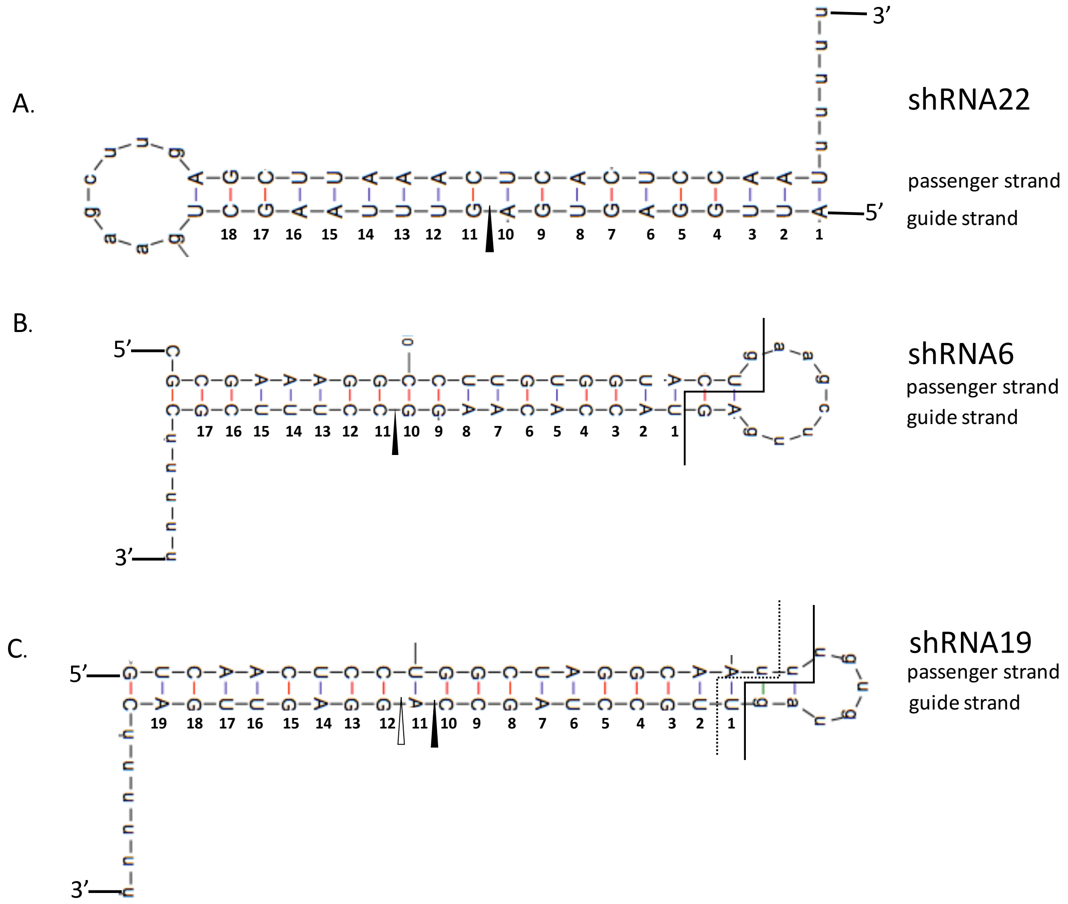


Figure 6.4 Illustration of shRNAs and RACE-Seq observed cleavage positions.

(A) illustrates the expected cleavage position for the shRNA22 analogue targeting HCV. Regardless of Dicer processing imprecision, due to the orientation of the guide strand, the expected RISC cleavage site is expected to remain unaltered. (B) indicates the expected cleavage position for shRNA6 (black triangle), with the numbers counting from the expected 5' end of the guide strand expected to be processed out of Dicer. For siRNA19, the black triangle represents the expected cleavage position of a shRNA19 analogue with a collapsed 5-nucleotide loop structure while the open triangle corresponds to a RACE-Seq confirmed cleavge position for a shRNA19 analogue with a 7-nucleotide loop structure. Lowercase letters = sequences that alogn to the loop the shRNA partner.

6.3 Discussion

For two groups of independently prepared and sequenced RACE-Seq samples, the total number of reads per sample generated on the Ion Torrent PGM was sufficient to identify the expected RISC cleavage site and in addition, proved sufficient to derive new insights from the RACE-Seq datasets. Pre-filtering of datasets prior to mapping of reads to the reference gene retained a sufficient amount of data for the read alignment and counting of 5' ends. Variability in the number of reads obtained for each sample did not impair identifying the expected cleavage site. A number of additional 5' peaks were observed around the expected peak site and within the expected siRNA hybridization site. Some of these were designated as degradation products, but in some instances, these additional peaks could be correlated to novel siRNA guide strands.

In this chapter, the RACE-Seq data was presented at the linear scale as well as percentage of aligned 5' ends. Graphically, the data is presented as bar graphs. RACE-Seq remains underutilized for confirming RISC cleavage, a small number of publications have with varied levels of success, applied the method. In most cases, the data is presented as percentage of aligned reads, without further reporting of total read number (Tabernero et al., 2013; Ganesh et al., 2016) and is used exclusively to report the expected RISC cleaved site. All current published RACE-Seq datasets show additional 5' end peaks within the guide strand hybridization site and for those publications that show a wider view, RACE-Seq is shown to report 5' ends well beyond the guide strand hybridization site (Tabernero et al., 2013; Barve et al., 2015).

Currently, there are no defined parameters for designing and executing RACE-Seq protocols or analysing and reporting the data. The assay itself remains challenging, requiring multiple manual handling operations that risk cross contamination of samples. Certainly, improvements to the RACE-Seq assay that eliminate the risk of cross contamination would be highly relevant if the assay is to be applied to clinical samples. For example, using barcoded RNA adapters at the RNA adapter ligation stage, would ensure that each sample is uniquely coded from the first sample handling stage (Peng et al., 2015; Kou et al., 2016).

When viewing the RACE-Seq peak profiles, it is assumed that the relative amounts presented at each position reflect the cleaved/ captured products of the original samples. RNA-ligase is known to show bias in RNA-RNA adapter ligation (Hafner et al., 2011). T4 RNA ligase mediated adapter ligation has been shown to have preference for certain end sequences that allow stable secondary structures between RNA ends (Raabe et al., 2014; Sorefan et al., 2012). The impact of adapter ligation bias for identifying and interpreting diverse RACE-Seq cleavage products remains to be investigated.

The shRNA6 samples produced the lowest number of NGS reads in both of the sequencing runs, with 5,812 reads for siRNA6(1) and an improved 29,934 reads for siRNA6(2) (though this was still 5 times below the next lowest read count). The dose response assays for siRNA6 and shRNA6 was calculated to have the same EC50 value, but the knockdown assays for RACE-Seq used 0.5 nmol/l for shRNA6(1) and 1.0 nmol/l for shRNA6(2). It seems that the number of reads obtained for shRNA6(1) therefore reflects the much lower amount of 5’ cleavage product that would have been generated. The increased dose would have resulted in an increase in cleavage products and therefore an increase in the amount of 5’ ends available for capture in the RLM RACE assay. Thus, in this case, the lower read depth does not constitute a failed assay, but rather correlates to the biological activity.

In this chapter, experiment pairs were critical for deriving and confirming novel observations for siRNA and shRNA design and activity. Most noteworthy has been the RACE-Seq analysis and accompanying theory for multiple RNAi triggers generated from siRNA19. Although the pattern of clustered 5’ ends varied across three independent samples, a distinct pattern emerged. Although only a very limited number of studies have been conducted that include a 23-mer siRNA, they have been shown to be substrates of Dicer (Kim et al., 2005). Additionally, the unusual mismatched duplex structure at the’ end of siRNA19 may interact in altered conformations with Dicer, resulting in different cleavages and a mix of active siRNA19 molecules. The proposed stem lengths that would generate the four RACE-Seq peak incidence are still all within the size range for AGO2 interaction. This hypothesis ties in with the current dogma of RISC cleavage specificity and variability of siRNA products from Dicer cleavage and seems more plausible than the proposal raised in Denise et al., (2014) of alternative/or imprecise slicer activity for AGO2. Their RACE-Seq assay generated a very similar peak profile.

One of the main risks of RNAi is the potential for off-target effects. There has been a long held concern that imprecise Dicer processing could lead to off-target effects by generating undesirable active siRNA (Dallas et al., 2012; Langenberger et al., 2013). These studies were largely based on the expression of short hairpin Dicer substrates with the sequence of these hairpins influenced by alterations in transcription start positions (Gu et al., 2012; Denise et al., 2014). For delivered Dicer substrate RNAi triggers such as synthetic shRNAs or long siRNA duplexes, RACE-Seq analysis (with confirmation of trigger sequence) has the potential to now provide evidence for an on-target mechanism of action of diverse RNAi triggers processed by Dicer. Modification of these Dicer substrates can then be investigated at an early stage to improve Dicer processing and drive the development of potent, clinically relevant RNAi drugs with well characterised RNAi mechanisms of action.

In conclusion, experience from this RACE-Seq work proposes that multiple independent samples are of greater value than read depth when determining novel insights from RACE-Seq data. Thus, low throughput sequencing instruments such as the Ion Torrent PGM and Illumina MiSeq are sufficient for multiplexing RACE-Seq samples. Here, RACE-Seq sample analysis defines a new role fort his assay, beyond confirming an expected RISC cleavage behaviour, RACE-Seq can define the sequence of the active siRNA.

7 ADDITIONAL FINDINGS FROM RACE-SEQ ASSAYS AND RACE-SEQ DATA ANALYSIS

7.1 Introduction

Beyond identifying the genetic sequence of living organisms, NGS technologies are increasingly being employed, together with novel capture assays, to detect and measure biological events. These new assays are often accompanied by or generate a need for bioinformatics tools. While access to NGS technology has largely been overcome with numerous commercial enterprises accepting various sample types, NGS data analysis remains a bottleneck. Additionally, most of the software packages that have been developed for NGS are executed in the Unix/Linux environment, and while there are commercial packages such as the CLC Genomics Workbench, these tools may be costly. In some instances, researchers may not have access to a bioinformatics core or a dedicated bioinformatician and may even lack sufficient knowledge of NGS to identify the necessary tools for analysing their data.

Traditionally, the 5' RACE assay accompanied by Sanger sequencing has been used to confirm an expected outcome of AGO2-RISC cleavage activity. However, transitioning these assays to NGS offers the unique opportunity to continue to expand our understanding of the complexity of the RNAi pathway. Additionally, this is an opportune time for the uptake of the RACE-Seq assay within the RNAi field, as the recent news of a successful Phase III clinical trial is set to spur the RNAi therapeutic field to new successes in 2018.

To this end, critical evaluation of current practise, establishing validated RACE-Seq assay protocols and appropriate data analysis practises will be helpful in persuading the RNAi therapeutic community of the added value that an NGS approach can provide. This chapter focuses on how standard and non-standard practises may impact RACE-Seq data output.

7.2 Results

7.2.1 RLM RACE assays for non-target control samples

It is standard practice to include a negative control in scientific experiments. However, attempting to sequence, by NGS, the negative (non-transfected control) sample from RNAi-based knockdown experiments may prove tricky. In order to perform RACE-Seq on the non-transfected cells from a knockdown assay, the RNA extracted from these cells must undergo all of the same operations as the treated sample.

In this example, RLM RACE assays were carried out for non-transfected controls for site 6, site 19, site 22 and the scrambled control for the TTR assays using the reverse primers and amplification primers in the same manner as the treated samples. All the samples had failed to produce a specific band at the expected 5' RACE position (see Figure 4.5 and Figure 5.1 A, control lane of gels). In all cases the control 5' RACE PCR samples were progressed directly to library preparation by first cleaning the samples using magnetic beads and quantifying the purified DNA using the Qubit High Sensitivity DNA assay. Qubit quantification yielded 29 ng/µl, 23 ng/µl, 14.4 ng/µl and 7 ng/µl, and allowed for 493 ng, 391 ng, 245 ng and 119 ng of purified DNA to be used as input to the end repair reactions for control 6, control 19, control 22 and control TTR samples respectively. After adapter ligation, the reactions were cleaned with 0.8X bead volume and the cleaned DNA amplified in 7 cycles of PCR. Samples were again cleaned using 0.8X bead volume and the final cleaned NGS libraries eluted in nuclease-free water. Samples were quantified by the Qubit assay and then analysed on the Bioanalyzer.

Unfortunately, in examining the fragment profiles of the control NGS samples, all samples were found to have fragment sizes that went well above 300 bp (Figure 7.1). For 200 bp sequencing on the Ion Torrent PGM, the maximum recommended size for library input is around 300 bp. The site 6 control sample had fragment size range from 323 bp to >2,000 bp, (Figure 7.1 A) while control 19 had fragment sizes ranging from 230 bp to 1,505 bp (Figure 7.1 B). Control 22 and the TTR scrambled control also have fragment sizes well outside of the acceptable range for sequencing on the Ion Torrent PGM (Figure 7.1 C and D). Since all the library preparations were well outside of the recommended input size for

Ion Torrent sequencing, these samples were not suitable for sequencing on the Ion Torrent PGM.

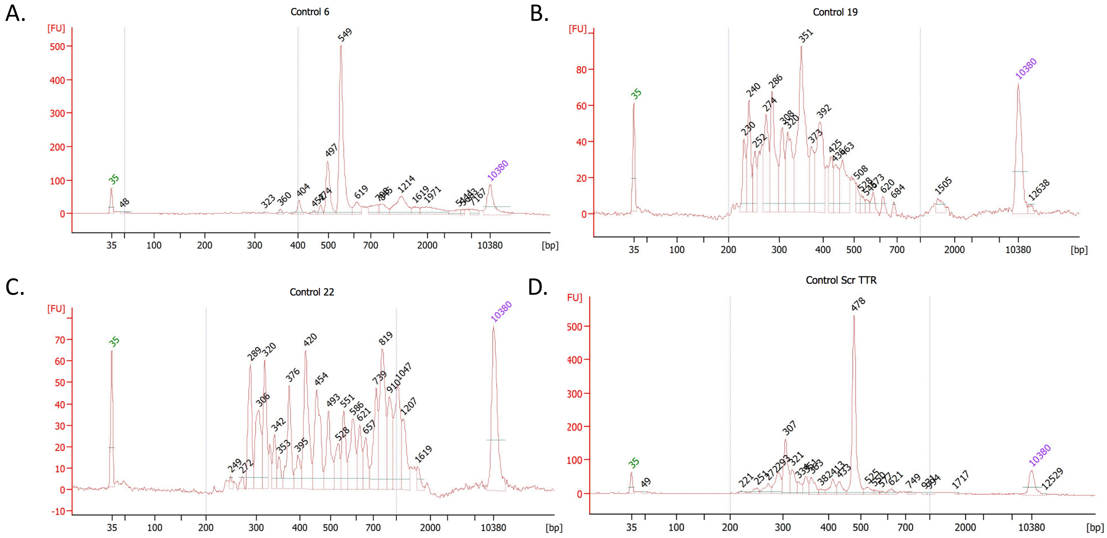


Figure 7.1 RACE-Seq library preparation for non-transfected control samples. (A) Control 6 sample, (B) Control 19 sample, (C) Control 22 sample and (D) Control TTR scrambled sample. All samples show DNA fragment sizes >300 bp in size and thus these libraries are not suitable for sequencing on the Ion Torrent PGM.

7.2.2 Read quality analysis of adapter+20-nt filtered datasets

A number of data pre-processing options for the long (>100 bp) amplicon RACE-Seq datasets was trialled before eventually opting for the adapter+20 nt approach. Since the reads had been trimmed to very short sequences (expected to be ~50 bases), quality filtering of reads was not implemented. Since quality filtering of reads removes low quality bases from the 3’ end of reads, the expectation was that the short reads would be of generally good quality. The FASTQC tool (Babraham Bioinformatics, https://www.bioinformatics.babraham.ac.uk/projects/download.html) which is a simple Java executed pipeline that allows read quality analysis to be carried out. In examining the per base read quality and the read length analysis outputs for each of the samples pre-and post-adapter+20 nt clipping of reads, some unexpected observations were made.

In comparing the profiles for read length, overall, a pattern regarding fragment size emerged. The majority of the time point 2 samples failed to be clipped to the ~50 bases stipulated by the adapter+20nt filtering criteria. The siRNA22(2) sample had at least 110,000 reads that were clipped to the short length, but more than 120,000 reads failed to be trimmed and were of length approximately 120 bases (Figure 7.2 B). Sample shRNA22(2) and sample siRNA6(2) were the only samples of time point 2 that were trimmed primarily to the expected ~50 bases (Figure 7.2 D and Figure 7.4 B) but the shRNA6(2), siRNA19(2) and shRNA19(2) samples each reported a second read length of approximately 110 bases (Figure 7.4 D, Figure 7.5 E and F respectively). Interestingly, the time point 2 samples were the ones that were size selected post library preparation and had been progressed directly to pooling for emulsion PCR without any cleanup. Further evidence that size selection post library preparation is likely to be the impacting activity, is that sample siRNA19(2*) and shRNA19(2*) were samples that had been size selected prior to library preparation but had undergone the NGS library preparation and sequencing operations of the time point 2 samples but these samples did not show inhibited length trimming when processing the data (Figure 7.5 C and D). This indicates that agarose gel size selection (including the E-gel size selection) imparted some ‘contaminating element’ to the DNA, and this had an impact on the sequences generated.

In analysing the per base quality of the processed RACE-Seq datasets, FASTQC analysis generated quite varied per base quality profiles for the siRNA22(1), siRNA22(2), shRNA22(1) and shRNA22(2) samples (Figure 7.2 A, B, C and D). Site 22 had already been identified as having greater sequence variability compared to site 6 and site 19 (Table 6.1 and 6.2; site 22 had higher percent unaligned filtered reads). For siRNA22(2), which had a mix of long and short reads, when the short reads and the longer reads were separated and the quality profiles analysed, the short reads length (~50 bases) of both size fractions showed improved per base quality profiles but the average per base quality scores across the length did vary (Figure 7.3).

It must also be remembered that, for this set of data, the GeneRacer RNA adapter sequence constitutes the first ~25 bases at the 5’ end of reads. Additionally, having reverse complemented reads that held the RNA adapter at the 3’ end, and pooled with the correct orientation adapter sequence, this may have placed possible poorer quality bases at the 5’ end of the dataset. This would account for the observed high variability of the per base quality scores observed in FASTQC profiles.

RACE-Seq samples siRNA6(1), siRNA6(2), shRNA6(1) and shRNA6(2) generally showed a much better per base quality profile compared to the site 22 samples. In general, the average per base quality values were above a score of 26 (Figures 7.4). Site 6 seemed to have less sequence variability as evidenced by >80% of filtered reads aligning to the HCV replicon reference (Table 6.1, Table 6.2). Again, with samples siRNA19(1), siRNA19(2*), siRNA19(2), shRNA19(1), shRNA19(2*) and shRNA19(2), FASTQC analysis identified a per base quality score generally above 26, particularly for nucleotide positions 1-50 (Figure 7.5).

In conclusion, using the freely accessible FASTQC tool has brought new insights to the impact of isolating DNA from agarose gels and the need for DNA clean-up after DNA isolation. RACE-Seq data analysis is likely to tolerate a lower threshold level of read quality as very short reads are being used in the final alignment and the data is aligned to a known sequence, however, it remains good practise to validate the quality of reads prior to implementing data analysis.

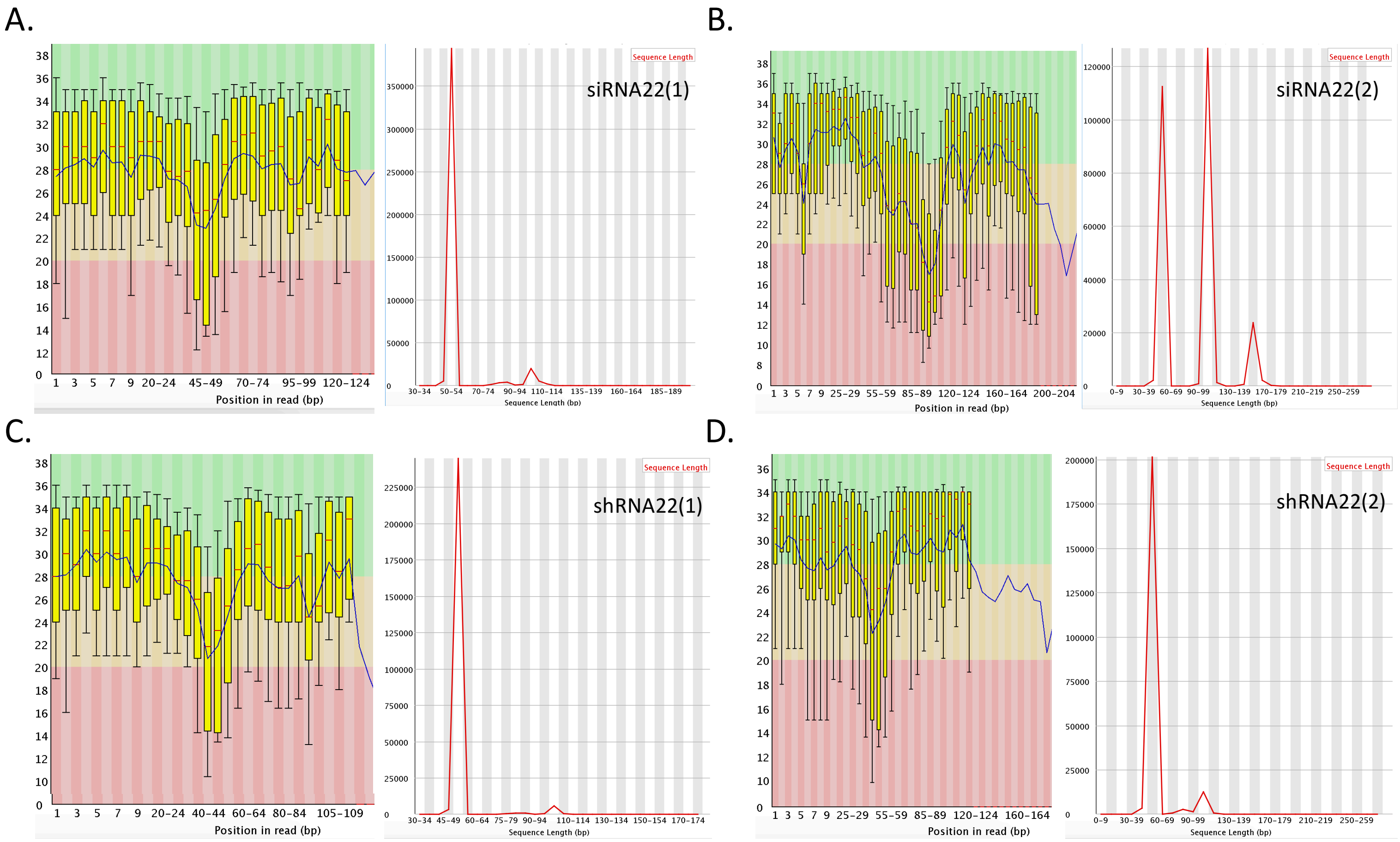


Figure 7.2 FASTQC analysis of RACE-Seq samples for site 22 samples for siRNA22 and shRNA22 samples

The per base quality plot and read length profiles (red peaks) for the filtered (adapter+20-nt) RACE-Seq datasets of (A) siRNA22(1), (B) siRNA22(2), (C) shRNA22(1) and (D) shRNA22(2) The average quality score = blue line

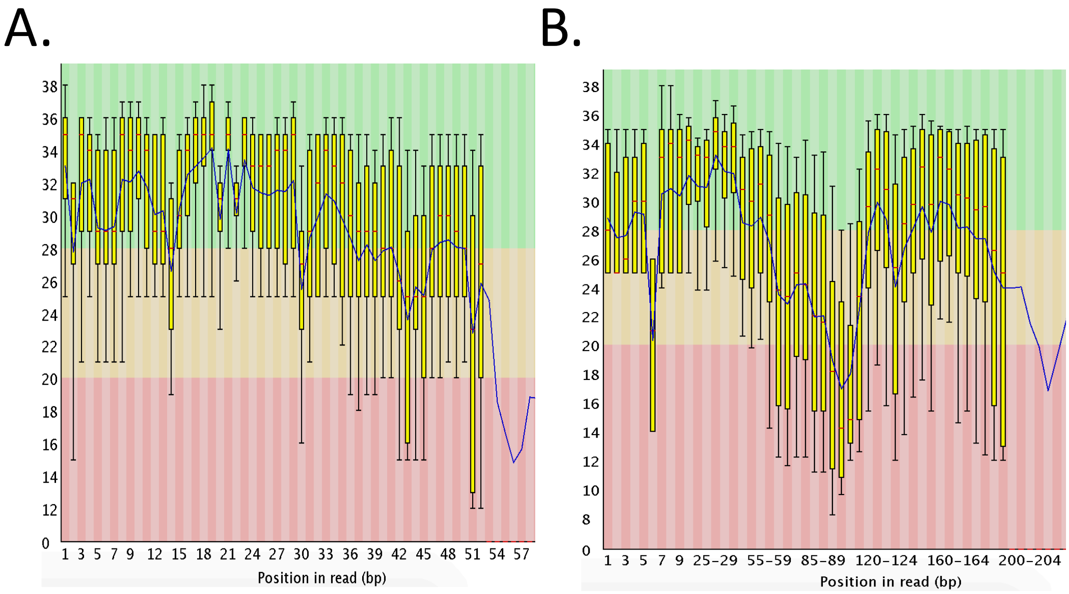


Figure 7.3 Per base quality plot for two different read lengths for RACE-Seq sample siRNA22(2).

For sample siRNA22(2), reads less than 60 bases were separated from reads greater than 60 bases from the filtered dataset (adapter+20-nt). The quality analysis for the short reads is presented in (A) and the quality analysis of the longs is presented in (B). The short reads has good quality scores, while that of the long reads varied.

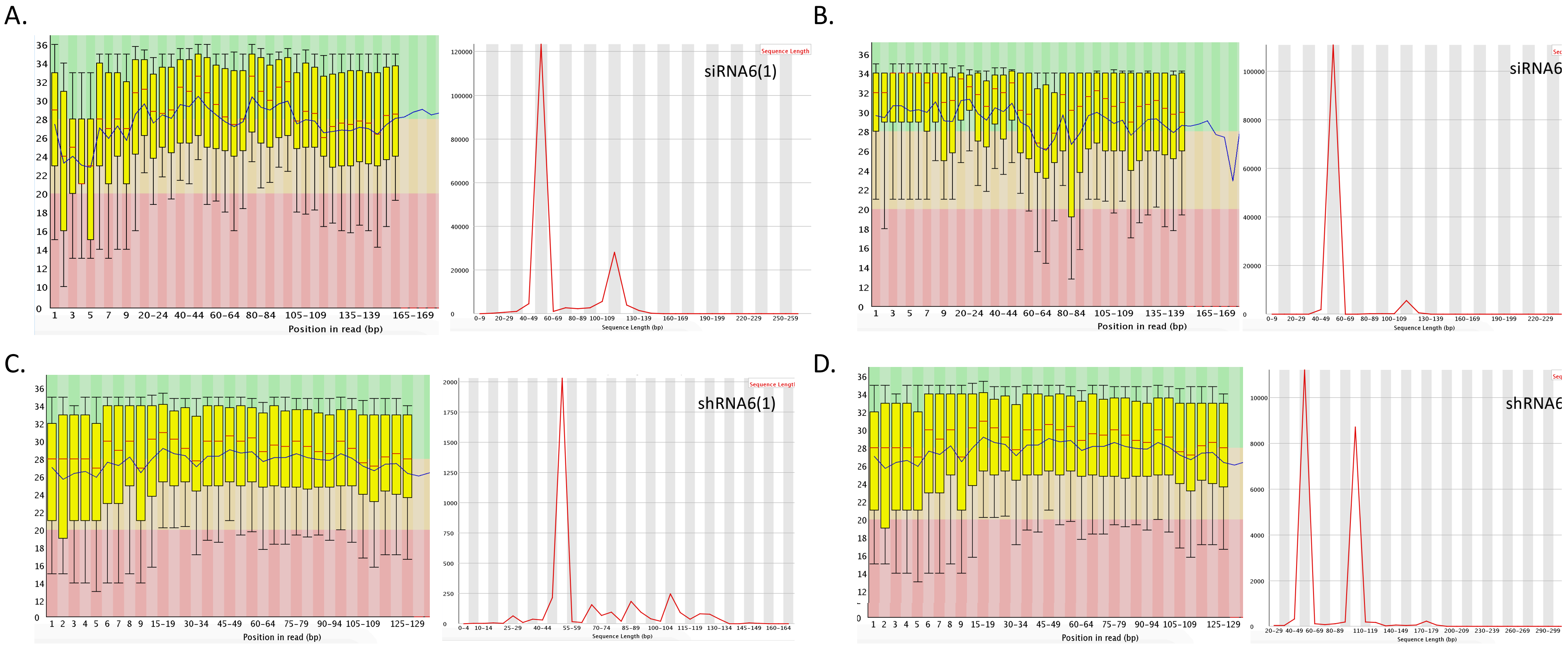
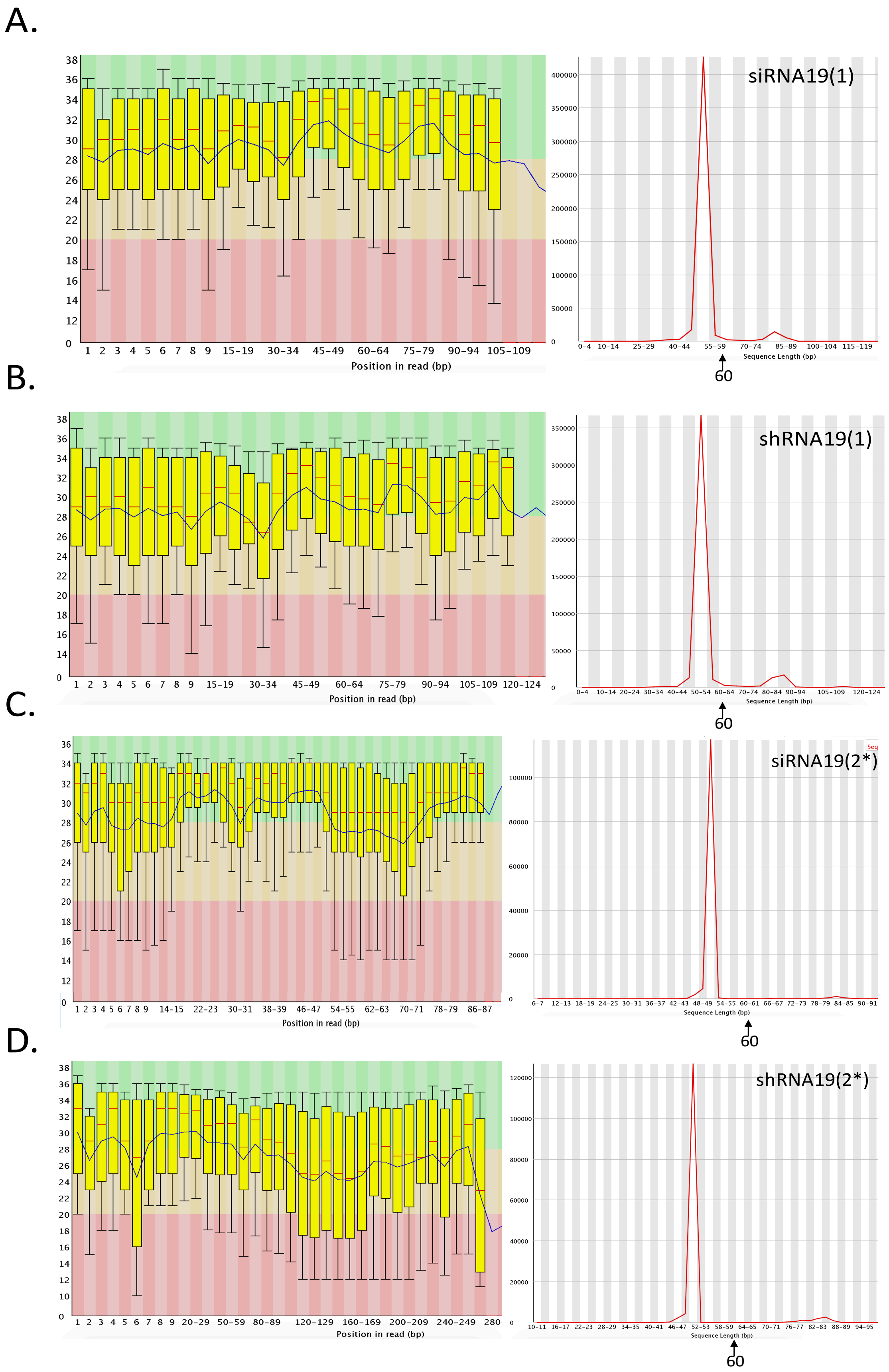


Figure 7.4 FASTQC analysis of RACE-Seq samples for site 6 samples for siRNA6 and shRNA6 samples.

The per base quality plot and the read length profiles for the filtered (adapter+20-nt) RACE-Seq datasets of (A) siRNA6(1), (B) siRNA6(2), (C) shRNA6 (1) and (D) shRNA6. The average quality score was above 26 (blue line) for all samples.


A.
Position in read (bp)
Sequence Length (bp)
siRNA19(1)
60
B.
Position in read (bp)
Sequence Length (bp)
shRNA19(1)
60
C.
Position in read (bp)
Sequence Length (bp)
siRNA19(2*)
60
D.
Position in read (bp)
Sequence Length (bp)
shRNA19(2*)
60

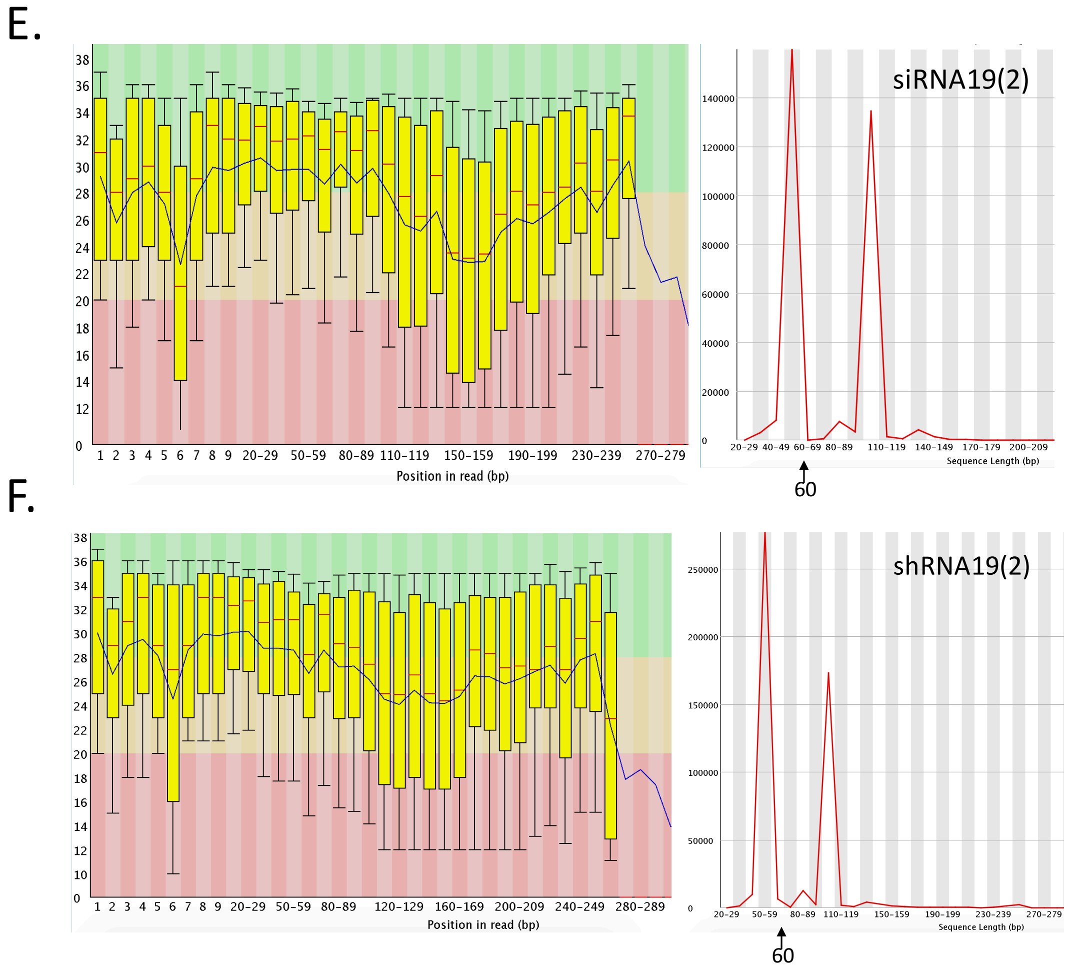


Figure 7.5 FASTQC analysis of RACE-Seq samples for site 19 samples for siRNA19 and shRNA19 samples.

Images on the left show the per base quality plot for each sample, and on the right the read length profiles for the filtered (adapter+20-nt) RACE-Seq datasets. The average quality score (blue line) was above 26 for all samples

7.2.3 Analysis and insights from failed TTR RACE-Seq datasets

PCR is widely used for targeted amplification of nucleic acid sequences and relies on well-designed primers, optimised PCR conditions and quality reagents. A set of 5' RACE samples from knockdown of *TTR* mRNA was analysed on 2% agarose gel. The expected DNA size bands of ~169 bp was not identified, but a smaller band was identified on the gel (Figure 7.6 A). Only a faint blurred region could be identified at 169 bp, but these samples were still progressed to library preparation. The 5' RACE PCR products for siRNA-TTR and DsiRNA-TTR were bead cleaned, end repaired and adapter ligated. After 7 cycles of PCR and a further round of bead cleaning, the samples were quantified, diluted and pooled with other NGS samples for emulsion PCR and sequencing on the Ion Torrent PGM. Data filtering and read alignment to the *TTR* mRNA reference found that this siRNA-TTR samples had only 15.7% of all reads that aligned to the *TTR* gene, while the DsiRNA-TTR sample had only 1.1% of total reads that aligned to the reference (Table 7.1). For siRNA-TTR the RACE-Seq dataset still reported the expected RISC cleaved position as the primary peak (Figure 7.6 B). Two other peaks were reported at T=480 and G=475 as12.1% and 24.6% (Table 7.2). On closer inspections, it was noteworthy that these positions corresponded to the expected cleavage products of DsiRNA-TTR. It may be that the siRNA-TTR sample was cross contaminated with the DsiRNA-TTR sample prior to NGS library preparation.

For the DsiRNA-TTR sample, although the expected RISC cleaved product was reported, it was not reported as the primary peak (Figure 7.6 C) and was reported as only 4.3% of aligned reads. Two other 5' peaks were identified. One peak corresponded to a cleaved product that indicated that a non-Dicer processed guide strand was bound to AGO2, as reported in Chapter 6. However, the peak at G=473 was reported as 45.7% of aligned reads. In this poorly executed RACE-Seq assay, this peak may falsely gain attention as a possible RISC derived product.

While these two examples have been pointedly examined as poor data it is noteworthy to observe that, if only the percent of 5' aligned reads were to be reported (Table 7.2), it would be much more challenging to determine that the expected cleavage positions were derived from a poor dataset. The siRNA-TTR sample reported the expected cleavage position as 51.2% of all aligned reads, and the DsiRNA-TTR sample reported the expected cleavage position as 10.4% aligned reads.

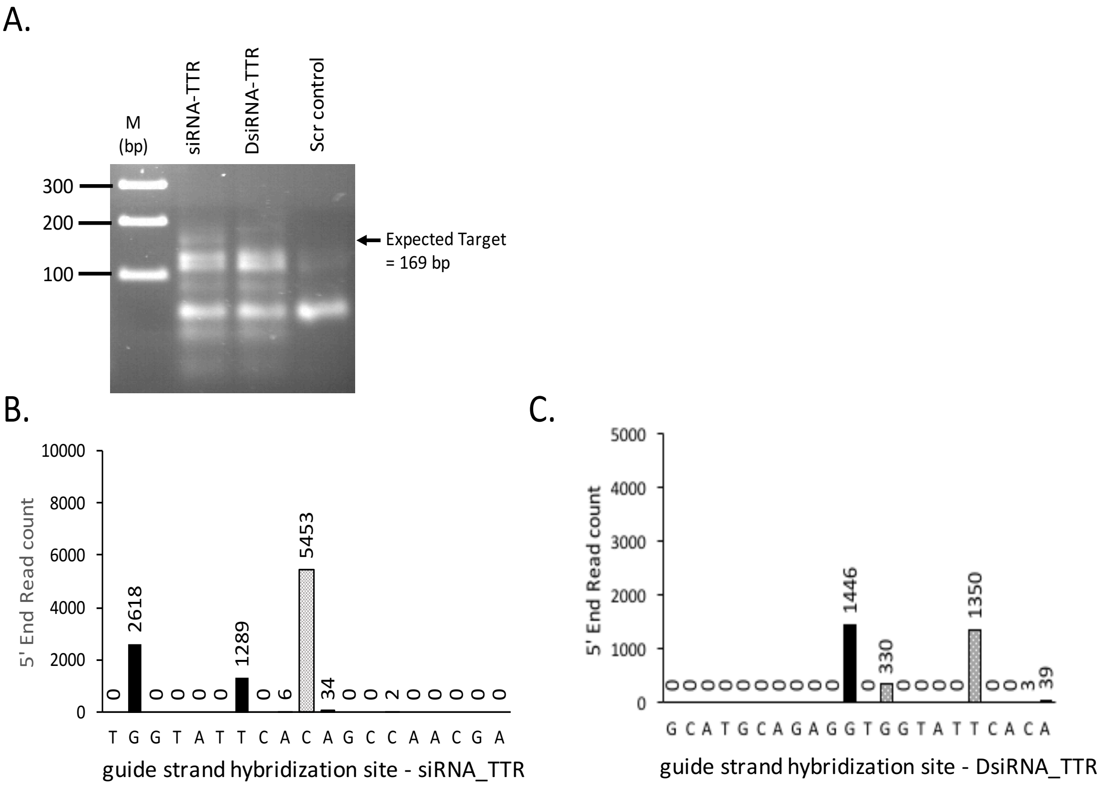


Figure 7.6 Analysis of failed TTR RACE-Seq sample

(A) Agarose gel analysis of failed sample showing prominent bands for siRNA-TTR and DsiRNA-TTR smaller than the expected cleavage site and no 5' RACE PCR band for the scrambled control sample. (B) RACE-Seq result for siRNA-TTR sample showing the expected cleavage position as the primary peak, as well as two additional prominent 5' end peaks. (C) RACE-Seq result for DsiRNA-TTR sample where the expected cleavage position is under reported.

Table 7.1 RACE-Seq read analysis of failed TTR sample

							Reads that align to target	
	Total Reads	Adapter +20nt	Reads with adapter	<10 bp	Total aligned reads	Unaligned reads	% of total reads	% of reads with adapter
siRNA-TTR(F)	68,102	11,939	11,318	0	10,658	660	15.7	94.2
DsiRNA-TTR(F)	282,622	3,716	3,346	1	3,166	179	1.1	94.6

Table 7.2 Table indicating the percentage of aligned reads for siRNA-TTR and DsiRNA-TTR failed RACE-Seq samples.

Hybridized guide strand position	Nucleotide	siRNA-TTR (% aligned)		Hybridized guide strand position	Nucleotide	DsiRNA-TTR (% aligned)
474	T	0.0	3’	463	G	0.0
475	G	24.6		464	C	0.0
476	G	0.0		465	A	0.0
477	T	0.0		466	T	0.0
478	A	0.0		467	G	0.0
479	T	0.0		468	C	0.0
480	T	12.1		469	A	0.0
481	C	0.0		470	G	0.0
482	A	0.1		471	A	0.0
483	C	51.2		472	G	0.0
484	A	0.3		473	G	45.7
485	G	0.0		474	T	0.0
486	C	0.0		475	G	10.4
487	C	0.0		476	G	0.0
488	A	0.0		477	T	0.0
489	A	0.0		478	A	0.0
490	C	0.0		479	T	0.0
491	G	0.0		480	T	42.6
492	A	0.0		481	C	0.0
				482	A	0.0
				483	C	0.1
				484	A	0.0
				485	G	0.0
				486	C	0.0
				487	C	0.0
				488	A	0.0
			5’	489	A	0.0

7.3 Discussion

In order for RACE-Seq to be used beyond reporting an expected outcome, numerous challenges in assay design, sample validation, NGS preparation, data analysis and data presentation are yet to be overcome. The four RACE-Seq assays currently in publication shed some light on these challenges (as discussed in Chapter 5). The challenges may be split into three; (1) preparing the RLM RACE assays (from RNA to PCR amplicon), (2) preparing the NGS libraries and (3) data analysis. However, since these three operations are in most instances likely to be carried out by three separate individuals/organisations, particularly in industry or large research institutes, and therefore, assessing and/or mitigating the impacts of the different operations may be more challenging.

Interestingly, it was found here, that certain sample preparation operations can impact data analysis in an unexpected manner. It seems that samples that had been size selected after library preparation had some inhibitory characteristic that prevented some of the NGS reads from being clipped to shorter reads by the fast-clipper programme from the FASTX toolkit. Access to such simple to use but powerful tools is vital in being able to analyse datasets prior to alignment. Since the 5' RACE assay is not specific for RISC cleaved products, fragmented target RNA derived by other means can be captured and amplified in the 5' RACE assay. Additionally, reverse transcription can generate non-target templates that hold the RNA adapter. Thus, 5' RACE is a 'dirty' assay, with spurious bands or background amplification often observed on agarose gels (Soutschek et al., 2004; Judge et al., 2009; Davis et al., 2010). For NGS, sample preparation methods that eliminate such material from the final library preparation help to enhance the amount of relevant data.

In conclusion, this analysis of failed RACE-Seq analysis highlights the need to validate each stage of the RACE-Seq assay and also highlights that even poorly conducted RACE-Seq assays, may provide sufficient data for confirming an expected proof of MOA for a given RNAi bioactive. In this light, it becomes imperative that a mechanism for qualifying good vs poor RACE-Seq assays is derived if RACE-Seq is to be utilised as an unbiased means for defining the active guide strand sequence based on an observed 5’ RACE-Seq peak.

8 CONCLUSIONS AND FUTURE WORK

RNAi-based therapeutics represent an innovative class of medicines that harness a natural pathway to target and cut disease relevant mRNA. Many of the initial difficulties that related to toxic side effects, poor efficacy and drug delivery have now been overcome and resulted in a resurgence in RNAi-based clinical trials from 2012, such that current development of clinically relevant drugs is based on over 15 years of technology and research development. An important part of RNAi research has been to understand the rules governing small RNA interactions with two important components within the RNAi pathway. These two components are Dicer and AGO2. Much of the research investigating and characterising the catalytic activity of Dicer and AGO2 substrates was undertaken from 2001 to 2006. Assays such as *in vitro* ^{32}P-radiolabeled synthetic mRNA cleavage and analysis on agarose gels (Elbashir et al., 2001; Schwarz et al., 2004; Rand et al., 2005b), northern blotting (Gu et al., 2012) and *in vitro* dicing assays analysed by gel electrophoresis and Electro-spray ionization liquid chromatography mass-spectroscopy (Kim et al., 2005; Rose et al., 2005) were used. The first publications using the 5' RACE assay to confirm the RISC-mediated cleavage product also began to emerge at this time (Soutschek et al., 2004; Zimmermann et al., 2006), however, the expectation for a particular outcome based on the design of the RNAi bioactive had already been established using the assays mentioned.

The 5' RACE assay with Sanger sequencing quickly arose as the gold standard for confirming direct RNAi-induced activity, with two rounds of PCR being performed to enrich for the desired target sequence. The second round PCR often employed a diluted volume of the first round PCR reaction (Rao et al., 2010), potentially selecting for only the most prominent RISC cleaved product, and therefore masking the presence of other potentially real RISC-derived products. Additionally, reporting of 5' RACE Sanger sequencing results has been limited to either a simple statement that the expected cleaved sequence was identified after sequencing a few clones or presenting an example sequencing chromatogram.

This thesis was concerned with designing and conducting RACE-Seq assays for the Ion Torrent PGM. The main aims of the project were to investigate robust methods for preparing NGS libraries for 5' RACE amplicons, to implement a user-friendly data analysis pipeline and to determine the potential for RACE-Seq to provide information regarding siRNA-induced AGO-driven MOA beyond the expected behaviour. The project succeeded in all three outcomes, but also raised a number of challenges for advancing the uptake of RACE-Seq assays within the RNAi field.

Some of the challenges in performing the assay include long incubation stages and substantial manual handling when processing the samples. These activities greatly reduce the number of samples that can be processed per day, or as a single batch. Additionally, manual handling operations increase the risk of cross-contamination as identified in at least two of the RACE-Seq samples presented in this thesis. Other than manual handling error, contamination may arise from repeat amplification of a particular sequence which leads to accumulation of amplification products in the lab environment. One proposal is to utilise a set of validated, barcoded RNA adapters, therefore tagging each RNA sample at the very first manipulation of the RNA.

In order for the RACE-Seq assay to be used to provide new information and deliver new insights regarding Dicer substrate impacts and RISC activity, more quantitative information regarding key aspects of the assay is required. T4 RNA ligase mediated bias has been extensively researched for small RNA NGS preparation (Hafner et al., 2011; Raabe et al., 2014). In particular, the work of Sorefan, *et al.*, (2012) comprehensively investigated the ligation preference T4 Rnl1. Although RACE-Seq has much lower diversity of target sequence compared to preparing a microRNA library, looking forward, an increase in the uptake of the assay would generate 'sequence diversity' and thus T4 RNA ligase-mediated bias may well impact RACE-Seq assays.

In this project, agarose gels were used to size select for NGS libraries. Exclusion of low molecular weight bands proved successful but exclusion of high molecular weight unwanted fragments was less successful. Although a number of electronic fragment separation systems are available, such equipment may not be readily available in all labs. The E-gel system is much more economical. However, this work identified that when the E-gel system was used to size select NGS libraries which were then diluted and progressed directly to emulsion PCR, it generated some contaminants that impacted the final datasets. When the PCR samples were size selected prior to library preparation, the bead clean operations are likely to have removed the contaminates, and as such, no inhibitory effects in trimming was observed for these samples. One of the difficulties in size selecting fragments prior to library preparation is that the amount of sample eluted may not be sufficient for NGS library preparation. This may be particularly impactful if using a commercial sequencing facility which would require a particular amount of high quality DNA. Alternative low cost size selection methods such as purification by denaturing polyacrylamide gel electrophoresis using the crush and soak extraction method (Alon et al., 2011) may be more successful in obtaining a discreet fragment size. Noteworthy is that the method used ethanol precipitation post gel extraction and highlighted that the use of Glycoblue during DNA precipitation is likely to interfere with Qubit-based quantification.

One of the biggest barriers to the uptake of the RACE-Seq assay is the lack of bioinformatics tools for analysing the data. We developed a simple to execute pipeline to analyse our RACE-Seq data, but some knowledge of Unix is required to execute the pipeline. The stand-alone pre-filtering command line code allows simple copy paste utility.

Moving forward, the novel insights regarding an on-target MOA for multiple guide strands derived from a single siRNA, Dicer substrate siRNA or shRNA needs to be investigated further.

RACE-Seq analysis may prove to be a useful tool for looking back at previous investigations of siRNA or shRNA design strategies. For example, overlaying RACE-Seq to an investigation of shRNA loop structure or sequence selection or positioning of a mismatch in the stem region has the potential to further enhance our understanding of Dicer and AGO2 processing preferences at the biological level. RACE-Seq also has the potential to be integrated into the design and validation of therapeutic oligos pre- and post clinical trial development.

REFERENCES

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P. and Axtell, M.J. (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current biology : CB*. 18 (10), 758–62.

Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D.C., Seidman, J.G., Church, G.M. and Eisenberg, E. (2011) Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome research*. 21 (9), 1506–11.

Amarzguioui, M., Lundberg, P., Cantin, E., Hagstrom, J., Behlke, M.A. and Rossi, J.J. (2006) Rational design and in vitro and in vivo delivery of Dicer substrate siRNA. *Nature Protocols*. 1 (2), 508–517.

Amarzguioui, M. and Rossi, J.J. (2008) '*Principles of Dicer Substrate (D-siRNA) Design and Function*', in [Online]. Humana Press. pp. 3–10.

Ansorge, W.J. (2009) Next-generation DNA sequencing techniques. *New Biotechnology*. 25 (4), 195–203.

Arezi, B., Xing, W., Sorge, J.A. and Hogrefe, H.H. (2003) Amplification efficiency of thermostable DNA polymerases. *Analytical biochemistry*. 321 (2), 226–35.

Barve, M., Wang, Z., Kumar, P., Jay, C.M., Luo, X., Bedell, C., Mennel, R.G., Wallraven, G., Brunicardi, F.C., Senzer, N., Nemunaitis, J. and Rao, D.D. (2015) Phase I Trial of bi-shRNA STMN1 BIV in Refractory Cancer. *Molecular therapy : the journal of the American Society of Gene Therapy*.

Baumberger, N. and Baulcombe, D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America*. 102 (33), 11928–33.

Benson, M.D., Kluve-Beckerman, B., Zeldenrust, S.R., Siesky, A.M., Bodenmiller, D.M., Showalter, A.D. and Sloop, K.W. (2006) Targeted suppression of an amyloidogenic transthyretin with antisense oligonucleotides. *Muscle & Nerve*. 33 (5), 609–618.

Berke, J.M., Vijgen, L., Lachau-Durand, S., Powdrill, M.H., Rawe, S., Sjuvarsson, E., Eriksson, S., Götte, M., Fransen, E., Dehertogh, P., Van den Eynde, C., Leclercq, L., Jonckers, T.H.M., Raboisson, P., Nilsson, M., Samuelsson, B., Rosenquist, Å., Fanning, G.C. and Lin, T.-I. (2011) Antiviral activity and mode of action of

TMC647078, a novel nucleoside inhibitor of the hepatitis C virus NS5B polymerase. *Antimicrobial agents and chemotherapy*. 55 (8), 3812–20.

Biegel, J.M. and Pager, C.T. (2016) Hepatitis C Virus Exploitation of Processing Bodies. *Journal of virology*. 90 (10), 4860–3.

Blight, K.J., McKeating, J.A., Marcotrigiano, J. and Rice, C.M. (2003) Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *Journal of virology*. 77 (5), 3181–90.

Blight, K.J., McKeating, J.A. and Rice, C.M. (2002) Highly Permissive Cell Lines for Subgenomic and Genomic Hepatitis C Virus RNA Replication. *Journal of Virology*. 76 (24), 13001–13014.

Borchert, G.M., Lanier, W. and Davidson, B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology*. 13 (12), 1097–1101.

Boudreau, R.L., Monteys, A.M. and Davidson, B.L. (2008) Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. *RNA*. 14 (9), 1834–1844.

Bracken, C.P., Szubert, J.M., Mercer, T.R., Dinger, M.E., Thomson, D.W., Mattick, J.S., Michael, M.Z. and Goodall, G.J. (2011) Global analysis of the mammalian RNA degradome reveals widespread miRNA-dependent and miRNA-independent endonucleolytic cleavage. *Nucleic acids research*. 39 (13), 5658–68.

Bronner, I.F., Quail, M.A., Turner, D.J., Swerdlow, H., Bronner, I.F., Quail, M.A., Turner, D.J. and Swerdlow, H. (2013) 'Improved Protocols for Illumina Sequencing', in *Current Protocols in Human Genetics*. [Online]. Hoboken, NJ, USA: John Wiley & Sons, Inc. p. 18.2.1-18.2.42.

Brown, N.P., Leroy, C. and Sander, C. (1998) MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics (Oxford, England)*. 14 (4), 380–1.

Butt, S., Idrees, M., Rehman, I., Ali, L., Hussain, A., Ali, M., Ahmed, N., Saleem, S. and Fayyaz, M. (2011) Establishment of stable Huh-7 cell lines expressing various hepatitis C virus genotype 3a protein: an in-vitro testing system for novel anti-HCV drugs. *Genetic Vaccines and Therapy*. 9 (1), 12.

Cai, X., Hagedorn, C.H. and Cullen, B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA (New York, N.Y.)*. 10 (12), 1957–66.

Cai, Y., Yu, X., Hu, S. and Yu, J. (2009) A Brief Review on the Mechanisms of miRNA Regulation. Genomics, Proteomics and Bioinformatics 7 (4) p.147–154.

Chang, B., Lee, C.H., Lee, J.H. and Lee, S.-W. (2010) Comparative analysis of intracellular inhibition of hepatitis C virus replication by small interfering RNAs. *Biotechnology letters*. 32 (9), 1231–7.

Che, Y., Ye, F., Xu, R., Qing, H., Wang, X., Yin, F., Cui, M., Burstein, D., Jiang, B. and Zhang, D.Y. (2012) Co-expression of XIAP and cyclin D1 complex correlates with a poor prognosis in patients with hepatocellular carcinoma. *The American journal of pathology*. 180 (5), 1798–807.

Chen, G., Kronenberger, P., Teugels, E. and De Grève, J. (2011) Influence of RT-qPCR primer position on EGFR interference efficacy in lung cancer cells. *Biological procedures online*. 131.

Chen, Z., Li, H., Ren, H., Hu, P., Lavanchy, D., Manns, M.P., Hadziyannis, S.J., Hunt, D., Pockros, P., Poordad, F., Lok, A.S., Osinusi, A., Zeuzem, S., Dahl, G., Sandstrom, A., Akerblom, E., Danielson, U.H., Flint, M., Cubero, M., Lu, L., Mo, H., Pilot-Matias, T.J., Molla, A., Lenz, O., Omata, M., Wen, C., Kuntzen, T., Hongmei, D.B.P.M., Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., Andino, R., Paolucci, S., Yoshimi, S., Lenz, O., Bartels, D.J., Lam, A.M., Manns, M.P., Cornberg, M., Asselah, T., Marcellin, P., Welsch, C., Tong, X., Tavis, J.E., Lin, C., Lin, C., Lenz, O., Lam, A.M., Karino, Y., Wang, C., Susser, S., Lagace, L., Kieffer, T.L., Sarrazin, C., Jiang, M., Fridell, R.A., Wong, K.A., Krishnan, P., Pilot-Matias, T., Kati, W., Lontok, E. and Sarrazin, C. (2016) Global prevalence of pre-existing HCV variants resistant to direct-acting antiviral agents (DAAs): mining the GenBank HCV genome data. *Scientific Reports*. 620310.

Chuang, C.F. and Meyerowitz, E.M. (2000) Specific and heritable genetic interference by double-stranded RNA in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*. 97 (9), 4985–90.

Clark, K.L., Hughes, S.A., Bulsara, P., Coates, J., Moores, K., Parry, J., Carr, M., Mayer, R.J., Wilson, P., Gruenloh, C., Levin, D., Darton, J., Weber, W.-M., Sobczak, K., Gill, D.R., Hyde, S.C., Davies, L.A., Pringle, I.A., Sumner-Jones, S.G., Jadhav, V., Jamison, S., Strapps, W.R., Pickering, V. and Edbrooke, M.R. (2013) Pharmacological Characterization of a Novel ENaCα siRNA (GSK2225745) With Potential for the Treatment of Cystic Fibrosis. *Molecular* *therapy. Nucleic acids*. 2e65.

Coelho, T., Adams, D., Silva, A., Lozeron, P., Hawkins, P.N., Mant, T., Perez, J., Chiesa, J., Warrington, S., Tranter, E., Munisamy, M., Falzone, R., Harrop, J., Cehelsky, J., Bettencourt, B.R., Geissler, M., Butler, J.S., Sehgal, A., Meyers, R.E., Chen, Q., Borland, T., Hutabarat, R.M., Clausen, V.A., Alvarez, R., Fitzgerald, K., Gamba-Vitalo, C., Nochur, S. V., Vaishnaw, A.K., Sah, D.W.Y., Gollob, J.A. and Suhr, O.B. (2013) Safety and Efficacy of RNAi Therapy for Transthyretin Amyloidosis. *New England Journal of Medicine*. 369 (9), 819–829.

Costa, R.H., Van Dyke, T.A., Yan, C., Kuo, F., Darnell, J.E. and Jr (1990) Similarities in transthyretin gene expression and differences in transcription factors: liver and yolk sac compared to choroid plexus. *Proceedings of the National Academy of Sciences of the United States of America*. 87 (17), 6589–93.

Dallas, A., Ilves, H., Ge, Q., Kumar, P., Shorenstein, J., Kazakov, S.A., Cuellar, T.L., McManus, M.T., Behlke, M.A. and Johnston, B.H. (2012) Right- and left-loop short shRNAs have distinct and unusual mechanisms of gene silencing. *Nucleic acids research*. 40 (18), 9255–71.

Davis, M.E., Zuckerman, J.E., Choi, C.H.J., Seligson, D., Tolcher, A., Alabi, C.A., Yen, Y., Heidel, J.D. and Ribas, A. (2010) Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature*. 464 (7291), 1067–70.

Denise, H., Moschos, S.A., Sidders, B., Burden, F., Perkins, H., Carter, N., Stroud, T., Kennedy, M., Fancy, S.-A., Lapthorn, C., Lavender, H., Kinloch, R., Suhy, D. and Corbau, R. (2014) Deep Sequencing Insights in Therapeutic shRNA Processing and siRNA Target Cleavage Precision. *Molecular therapy. Nucleic acids*. 3e145.

Devogelaere, B., Berke, J.M., Vijgen, L., Dehertogh, P., Fransen, E., Cleiren, E., van der Helm, L., Nyanguile, O., Tahri, A., Amssoms, K., Lenz, O., Cummings, M.D., Clayton, R.F., Vendeville, S., Raboisson, P., Simmen, K.A., Fanning, G.C. and Lin, T.-I. (2012) TMC647055, a potent nonnucleoside hepatitis C virus NS5B polymerase inhibitor with cross-genotypic coverage. *Antimicrobial agents and chemotherapy*. 56 (9), 4676–84.

Ding, Y., Chan, C.Y. and Lawrence, C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic acids research*. 32 (Web Server issue), W135-41.

Dubuisson, J. and Cosset, F.-L. (2014) Virology and cell biology of the hepatitis C virus life cycle – An update. *Journal of Hepatology*. 61 (1), S3–S13.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*. 411 (6836), 494–8.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T., Ambros, V., Bass, B., Bernstein, E., Caudy, A., Hammond, S., Hannon, G., Caplen, N., Parrish, S., Imani, F., Fire, A., Morgan, R., Carthew, R., Cerutti, L., Mian, N., Bateman, A., Clemens, J., Worby, C., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B., Dixon, J., Cogoni, C., Macino, G., Dalmay, T., Hamilton, A., Rudd, S., Angell, S., Baulcombe, D., Elbashir, S., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., Elbashir, S., Lendeckel, W., Tuschl, T., Fire, A., Fire, A., Xu, S., Montgomery, M., Kostas, S., Driver, S., Mello, C., Fraser, A., Kamath, R., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., Ahringer, J., Gönczy, P., Grishok, A., Hamilton, A., Baulcombe, D., Hammond, S., Bernstein, E., Beach, D., Hannon, G., Hammond, S., Boettcher, S., Caudy, A., Kobayashi, R., Hannon, G., Hammond, S., Caudy, A., Hannon, G., Hutvágner, G., Mlynarova, L., Nap, J., Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., Zamore, P., Jacobsen, S., Running, M., Meyerowitz, M., Knight, S., Bass, B., Lam, G., Thummel, C., Maeda, I., Kohara, Y., Yamamoto, M., Sugimoto, A., Matsuda, S., Ichigotani, Y., Okuda, T., Irimura, T., Nakatsugawa, S., Hamaguchi, M., Moss, E., Mourrain, P., Nykänen, A., Haley, B., Zamore, P., Olsen, P., Ambros, V., Palauqui, J., Elmayan, T., Pollien, J., Vaucheret, H., Parrish, S., Fleenor, J., Xu, S., Mello, C., Fire, A., Pasquinelli, A., Piano, F., Schetterdagger, A., Mangone, M., Stein, L., Kemphues, K., Ray, A., Lang, J., Golden, T., Ray, S., Reinhart, B., Slack, F., Basson, M., Pasquinelli, A., Bettinger, J., Rougvie, A., Horvitz, H., Ruvkun, G., Sharp, P., Smardon, A., Spoerke, J., Stacey, S., Klein, M., Mackin, N., Maine, E., Stark, G., Kerr, I., Williams, B., Silverman, R., Schreiber, R., Tuschl, T., Tuschl, T., Zamore, P., Lehmann, R., Bartel, D., Sharp, P., Ui-Tei, K., Zenno, S., Miyata, Y., Saigo, K., Voinnet, O., Voinnet, O., Vain, P., Angell, S., Baulcombe, D., Waterhouse, P., Wang, M., Lough, T., Wightman, B., Ha, I., Ruvkun, G., Yang, D., Lu, H., Erickson, J., Zamore, P., Tuschl, T., Sharp, P. and Bartel, D. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in

Drosophila melanogaster embryo lysate. *The EMBO journal*. 20 (23), 6877–88.

Endrullat, C., Glökler, J., Franke, P. and Frohme, M. (2016) Standardization and quality management in next-generation sequencing. *Applied & Translational Genomics*. 102–9.

Feng, Y., Zhang, X., Graves, P. and Zeng, Y. (2012) A comprehensive analysis of precursor microRNA cleavage by human Dicer. *RNA (New York, N.Y.)*. 18 (11), 2083–92.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. 391 (6669), 806–11.

Flores-Jasso, C.F., Arenas-Huertero, C., Reyes, J.L., Contreras-Cubas, C., Covarrubias, A. and Vaca, L. (2009) First step in pre-miRNAs processing by human Dicer. *Acta pharmacologica Sinica*. 30 (8), 1177–85.

Foster, D.J., Barros, S., Duncan, R., Shaikh, S., Cantley, W., Dell, A., Bulgakova, E., O'Shea, J., Taneja, N., Kuchimanchi, S., Sherrill, C.B., Akinc, A., Hinkle, G., Seila White, A.C., Pang, B., Charisse, K., Meyers, R., Manoharan, M. and Elbashir, S.M. (2012) Comprehensive evaluation of canonical versus Dicer-substrate siRNA in vitro and in vivo. *RNA (New York, N.Y.)*. 18 (3), 557–68.

Frank-Kamenetsky, M., Grefhorst, A., Anderson, N.N., Racie, T.S., Bramlage, B., Akinc, A., Butler, D., Charisse, K., Dorkin, R., Fan, Y., Gamba-Vitalo, C., Hadwiger, P., Jayaraman, M., John, M., Jayaprakash, K.N., Maier, M., Nechev, L., Rajeev, K.G., Read, T., Röhl, I., Soutschek, J., Tan, P., Wong, J., Wang, G., Zimmermann, T., de Fougerolles, A., Vornlocher, H.-P., Langer, R., Anderson, D.G., Manoharan, M., Koteliansky, V., Horton, J.D. and Fitzgerald, K. (2008) Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America*. 105 (33), 11915–20.

Frese, M., Schwärzle, V., Barth, K., Krieger, N., Lohmann, V., Mihm, S., Haller, O. and Bartenschlager, R. (2002) Interferon-γ inhibits replication of subgenomic and genomic hepatitis C virus RNAs. *Hepatology*. 35 (3), 694–703.

Frohman, M.A. (1994) On beyond classic RACE (rapid amplification of cDNA ends). *PCR methods and applications*. 4 (1), S40-58.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America*. 85 (23), 8998–9002.

Ganesh, S., Koser, M.L., Cyr, W.A., Chopda, G.R., Tao, J., Shui, X., Ying, B., Chen, D., Pandya, P., Chipumuro, E., Siddiquee, Z., Craig, K., Lai, C., Dudek, H., Monga, S.P., Wang, W., Brown, B.D. and Abrams, M.T. (2016) Direct Pharmacological Inhibition of b-Catenin by RNA Interference in Tumors of Diverse Origin. *Mol Cancer Ther*. 15 (9), 1–12.

German, M.A., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L.A., Nobuta, K., German, R., De Paoli, E., Lu, C., Schroth, G., Meyers, B.C. and Green, P.J. (2008) Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology*. 26 (8), 941–946.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*. 27 (1), 91–105.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N. and Kay, M.A. (2012) The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell*. 151 (4), 900–11.

Gu, T.J., Yi, X., Zhao, X.W., Zhao, Y. and Yin, J.Q. (2009) Alu-directed transcriptional regulation of some novel miRNAs. *BMC Genomics*. 10 (1), 563.

Guo, S., Kemphues, K.J., White, J., Thomson, N., Lim, L., Brennan, C.H., Sidebottom, C., Davison, M.D. and Scott, J. (1995) par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*. 81 (4), 611–20.

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J.T.G., Nusbaum, J.D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G. and Tuschl, T. (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA (New York, N.Y.)*. 17 (9), 1697–712.

Hagopian, J.C., Hamil, A.S., van den Berg, A., Meade, B.R., Eguchi, A., Palm-Apergi, C. and Dowdy, S.F. (2017) Induction of RNAi Responses by Short Left-Handed Hairpin RNAi Triggers. *Nucleic Acid Therapeutics*. 27 (5), 260–271.

Hamilton, A.J. and Baulcombe, D.C. (1999) A Species of Small Antisense RNA in

Posttranscriptional Gene Silencing in Plants. *Science*. 286 (5441), .

Harwig, A., Herrera-Carrillo, E., Jongejan, A., van Kampen, A.H. and Berkhout, B. (2015) Deep Sequence Analysis of AgoshRNA Processing Reveals 3’ A Addition and Trimming. *Molecular therapy. Nucleic acids*. 4 (7), e247.

Hayashi, Y., Mori, Y., Yamashita, S., Motoyama, K., Higashi, T., Jono, H., Ando, Y. and Arima, H. (2012) Potential Use of Lactosylated Dendrimer (G3)/α-Cyclodextrin Conjugates as Hepatocyte-Specific siRNA Carriers for the Treatment of Familial Amyloidotic Polyneuropathy. *Molecular Pharmaceutics*. 9 (6), 1645–1653.

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R. and Ordoukhanian, P. (2014) Library construction for next-generation sequencing: overviews and challenges. *BioTechniques*. 56 (2), 61–4, 66, 68, passim.

Heidersbach, A., Gaspar-Maia, A., McManus, M.T. and Ramalho-Santos, M. (2006) RNA interference in embryonic stem cells and the prospects for future therapies. *Gene Therapy*. 13 (6), 478–486.

Helvik, S.A., Snove, O. and Saetrom, P. (2007) Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics*. 23 (2), 142–149.

Herbert, M., Coppieters, N., Lasham, A., Cao, H. and Reid, G. (2011) The importance of RT-qPCR primer design for the detection of siRNA-mediated mRNA silencing. *BMC Research Notes*. 4 (1), 148.

Holmes, K., Williams, C.M., Chapman, E.A. and Cross, M.J. (2010) Detection of siRNA induced mRNA silencing by RT-qPCR: considerations for experimental design. *BMC research notes*. 353.

Hosoi, A., Su, Y., Torikai, M., Jono, H., Ishikawa, D., Soejima, K., Higuchi, H., Guo, J., Ueda, M., Suenaga, G., Motokawa, H., Ikeda, T., Senju, S., Nakashima, T. and Ando, Y. (2016) Novel Antibody for the Treatment of Transthyretin Amyloidosis. *The Journal of biological chemistry*. 291 (48), 25096–25105.

Ipsaro, J.J. and Joshua-Tor, L. (2015) From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nature Structural & Molecular Biology*. 22 (1), 20–28.

Jackson, A.L. and Linsley, P.S. (2010) Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature reviews. Drug discovery*. 9 (1), 57–67.

Janssen, H.L. a, Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patick, A.K., Chen, A., Zhou, Y., Persson, R., King, B.D., Kauppinen, S., Levin, A. a and Hodges, M.R. (2013) Treatment of HCV infection by targeting . *The New England journal of medicine*. 368 (18), 1685–94.

Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M. and Sarnow, P. (2005) Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA. *Science*. 309 (5740), .

Judge, A.D., Robbins, M., Tavakoli, I., Levi, J., Hu, L., Fronda, A., Ambegia, E., McClintock, K. and MacLachlan, I. (2009) Confirming the RNAi-mediated mechanism of action of siRNA-based cancer therapeutics in mice. *The Journal of clinical investigation*. 119 (3), 661–73.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & development*. 15 (20), 2654–9.

Kim, D.-H., Behlke, M.A., Rose, S.D., Chang, M.-S., Choi, S. and Rossi, J.J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nature Biotechnology*. 23 (2), 222–226.

Kim, Y., Kim, V.N., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A. and Nogales, E. (2012) MicroRNA factory: RISC assembly from precursor microRNAs. *Molecular cell*. 46 (4), 384–6.

Kohn, A.B., Moroz, T.P., Barnes, J.P., Netherton, M. and Moroz, L.L. (2013) Single-cell semiconductor sequencing. *Methods in molecular biology (Clifton, N.J.)*. 1048247–84.

Koller, E., Propp, S., Murray, H., Lima, W., Bhat, B., Prakash, T.P., Allerson, C.R., Swayze, E.E., Marcusson, E.G. and Dean, N.M. (2006) Competition for RISC binding predicts in vitro potency of siRNA. *Nucleic Acids Research*. 34 (16), 4467–4476.

Kou, R., Lam, H., Duan, H., Ye, L., Jongkam, N., Chen, W., Zhang, S. and Li, S. (2016) Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations. *PloS one*. 11 (1), e0146638.

Krieger, N., Lohmann, V. and Bartenschlager, R. (2001) Enhancement of hepatitis C virus RNA replication by cell culture-adaptive mutations. *Journal of virology*. 75 (10), 4614–24.

Krönke, J., Kittler, R., Buchholz, F., Windisch, M.P., Pietschmann, T., Bartenschlager, R. and Frese, M. (2004) Alternative approaches for efficient inhibition of hepatitis C virus RNA replication by small interfering RNAs. *Journal of virology*. 78 (7), 3436–46.

Langenberger, D., Çakir, M.V., Hoffmann, S. and Stadler, P.F. (2013) Dicer-processed small RNAs: rules and exceptions. *Journal of experimental zoology. Part B, Molecular and developmental evolution*. 320 (1), 35–46.

Lasham, A., Herbert, M., Coppieters 't Wallant, N., Patel, R., Feng, S., Eszes, M., Cao, H. and Reid, G. (2010) A rapid and sensitive method to detect siRNA-mediated mRNA cleavage in vivo using 5' RACE and a molecular beacon probe. *Nucleic acids research*. 38 (3), e19.

Lau, P.-W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B. and MacRae, I.J. (2012) The molecular architecture of human Dicer. *Nature structural & molecular biology*. 19 (4), 436–40.

Lavender, H., Brady, K., Burden, F., Delpuech-Adams, O., Denise, H., Palmer, A., Perkins, H., Savic, B., Scott, S., Smith-Burchnell, C., Troke, P., Wright, J.F., Suhy, D. and Corbau, R. (2012) In vitro characterization of the activity of PF-05095808, a novel biological agent for hepatitis C virus therapy. *Antimicrobial agents and chemotherapy*. 56 (3), 1364–75.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal*. 23 (20), 4051–60.

Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell 120 (1) p.15–20.

Li, Y., Yamane, D., Masaki, T. and Lemon, S.M. (2015) The yin and yang of hepatitis C: synthesis and decay of hepatitis C virus RNA. *Nature reviews. Microbiology*. 13 (9), 544–58.

Liang, X.-H., Hart, C.E. and Crooke, S.T. (2013) Transfection of siRNAs can alter miRNA levels and trigger non-specific protein degradation in mammalian cells. *Biochimica et Biophysica Acta*. 1829 (5), 455–68.

Lin, X., Morgan-Lappe, S., Huang, X., Li, L., Zakula, D.M., Vernetti, L.A., Fesik, S.W. and Shen, Y. (2007) 'Seed' analysis of off-target siRNAs reveals an essential role of Mcl-1 in resistance to the small-molecule Bcl-2/Bcl-XL inhibitor ABT-737. *Oncogene*. 26 (27), 3972–9.

Liu, Y.P., Karg, M., Harwig, A., Herrera-Carrillo, E., Jongejan, A., Kampen, A. van and Berkhout, B. (2015) *Mechanistic insights on the Dicer-independent AGO2-mediated processing of AgoshRNAs*.

Lohmann, V., Körner, F., Dobierzewska, A. and Bartenschlager, R. (2001) Mutations in hepatitis C virus RNAs conferring cell culture adaptation. *Journal of virology*. 75 (3), 1437–49.

Lohmann, V., Körner, F., Koch, J.-O., Herian, U., Theilmann, L. and Bartenschlager, R. (1999) Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line. *Science*. 285 (5424), .

Lu, L., Pilot-Matias, T.J., Stewart, K.D., Randolph, J.T., Pithawalla, R., He, W., Huang, P.P., Klein, L.L., Mo, H. and Molla, A. (2004) Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor in vitro. *Antimicrobial agents and chemotherapy*. 48 (6), 2260–6.

Luna, J.M., Scheel, T.K.H., Danino, T., Shaw, K.S., Mele, A., Fak, J.J., Nishiuchi, E., Takacs, C.N., Catanese, M.T., de Jong, Y.P., Jacobson, I.M., Rice, C.M. and Darnell, R.B. (2015) Hepatitis C virus RNA functionally sequesters miR-122. *Cell*. 160 (6), 1099–110.

Lund, E. and Dahlberg, J.E. (2006) Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor symposia on quantitative biology*. 7159–66.

Ma, E., Zhou, K., Kidwell, M.A. and Doudna, J.A. (2012) Coordinated activities of human dicer domains in regulatory RNA processing. *Journal of Molecular Biology*. 422 (4), 466–476.

Machlin, E.S., Sarnow, P. and Sagan, S.M. (2011) Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (8), 3193–8.

MacRae, I.J. (2006) Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*. 311 (5758), 195–198.

Malapelle, U., Vigliar, E., Sgariglia, R., Bellevicine, C., Colarossi, L., Vitale, D., Pallante, P. and Troncone, G. (2015) Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. *Journal of Clinical Pathology*. 68 (1), 64–68.

Martinez, J. and Tuschl, T. (2004) RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes & development*. 18 (9), 975–80.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., Zamore, P.D., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., Zamore, P.D., Gaul, U. and Rajewsky, N. (2005) Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell*. 123 (4), 607–620.

McAnuff, M.A., Rettig, G.R. and Rice, K.G. (2007) Potency of siRNA versus shRNA mediated knockdown in vivo. *Journal of pharmaceutical sciences*. 96 (11), 2922–30.

McIntyre, G. and Fanning, G. (2006) Design and cloning strategies for constructing shRNA expression vectors. *BMC Biotechnology*. 6 (1), 1.

McWilliam Leitch, E.C. and McLauchlan, J. (2013) Determining the cellular diversity of hepatitis C virus quasispecies by single-cell viral sequencing. *Journal of virology*. 87 (23), 12648–55.

Melamed, Z., Levy, A., Ashwal-Fluss, R., Lev-Maor, G., Mekahel, K., Atias, N., Gilad, S., Sharan, R., Levy, C., Kadener, S. and Ast, G. (2013) Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. *Molecular cell*. 50 (6), 869–81.

Merriman, B., R&D Team, I.T. and Rothberg, J.M. (2012) Progress in Ion Torrent semiconductor chip based sequencing. *ELECTROPHORESIS*. 33 (23), 3397–3417.

Nakanishi, K. (2016) Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley interdisciplinary reviews. RNA*. 7 (5), 637–60.

Neff, C.P., Zhou, J., Remling, L., Kuruvilla, J., Zhang, J., Li, H., Smith, D.D., Swiderski, P., Rossi, J.J. and Akkina, R. (2011) An aptamer-siRNA chimera suppresses HIV-1 viral loads and protects from helper CD4(+) T cell decline in humanized mice. *Science translational medicine*. 3 (66), 66ra6.

Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *Journal of Molecular Biology*. 302 (1), 205–217.

Orban, T.I. and Izaurralde, E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA (New York, N.Y.)*. 11 (4), 459–69.

Palha, J., Moreira, P., Olofsson, A., Lundgren, E. and Saraiva, M. (2002) Antibody recognition of amyloidogenic transthyretin variants in serum of patients with familial amyloidotic polyneuropathy. *Journal of the Peripheral Nervous System*. 7 (2), 134–134.

Pang, P.S., Pham, E.A., Elazar, M., Patel, S.G., Eckart, M.R. and Glenn, J.S. (2012) Structural map of a microRNA-122: hepatitis C virus complex. *Journal of virology*. 86 (2), 1250–4.

Park, C.-W., Cho, M.-C., Hwang, K., Ko, S.-Y., Oh, H.-B. and Lee, H.C. (2014) Comparison of Quasispecies Diversity of HCV between Chronic Hepatitis C and Hepatocellular Carcinoma by Ultradeep Pyrosequencing. *BioMed Research International*. 2014.

Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. and Kim, V.N. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*. 475 (7355), 201–5.

Patel, K., Kilfoil, G., Wyles, D.L., Naggie, S., Lawitz, E., Bradley, S., Lindell, P. and Suhy, D. (2016) 258. Phase I/IIa Study of TT-034, a DNA-Directed RNA Interference (ddRNAi) Agent Delivered as a Single Administration for the Treatment of Subjects with Chronic Hepatitis C Virus (HCV). *Molecular Therapy*. 24S102.

Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P. and Wang, Y. (2015) Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC genomics*. 16 (1), 589.

Raabe, C.A., Tang, T.-H., Brosius, J. and Rozhdestvensky, T.S. (2014) Biases in small RNA deep sequencing data. *Nucleic acids research*. 42 (3), 1414–26.

Ramalingam, P., Palanichamy, J.K., Singh, A., Das, P., Bhagat, M., Kassab, M.A., Sinha, S. and Chattopadhyay, P. (2014) Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA (New York, N.Y.)*. 20 (1), 76–87.

Rand, T.A., Petersen, S., Du, F. and Wang, X. (2005a) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*. 123 (4), 621–629.

Rand, T.A., Petersen, S., Du, F. and Wang, X. (2005b) Argonaute2 Cleaves the Anti-Guide Strand of siRNA during RISC Activation. *Cell*. 123 (4), 621–629.

Rao, D.D., Maples, P.B., Senzer, N., Kumar, P., Wang, Z., Pappen, B.O., Yu, Y., Haddock, C., Jay, C., Phadke, A.P., Chen, S., Kuhn, J., Dylewski, D., Scott, S., Monsma, D., Webb, C., Tong, A., Shanahan, D. and Nemunaitis, J. (2010) Enhanced target gene knockdown by a bifunctional shRNA: a novel approach of RNA interference. *Cancer gene therapy*. 17 (11), 780–91.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S. and Khvorova, A. (2004) Rational siRNA design for RNA interference. *Nature Biotechnology*. 22 (3), 326–330.

Riss, T.L., Moravec, R.A., Niles, A.L., Duellman, S., Benink, H.A., Worzella, T.J. and Minor, L. (2004) *Cell Viability Assays*. Eli Lilly & Company and the National Center for Advancing Translational Sciences.

Robbins, M., Judge, A. and MacLachlan, I. (2009) siRNA and innate immunity. *Oligonucleotides*. 19 (2), 89–102.

Rodrigue, S., Materna, A.C., Timberlake, S.C., Blackburn, M.C., Malmstrom, R.R., Alm, E.J. and Chisholm, S.W. (2010) Unlocking short read sequencing for metagenomics. *PloS one*. 5 (7), e11840.

Rose, S.D., Kim, D.-H., Amarzguioui, M., Heidel, J.D., Collingwood, M.A., Davis, M.E., Rossi, J.J. and Behlke, M.A. (2005) Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic acids research*. 33 (13), 4140–56.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F., Marran, D., Myers, J.W., Davidson, J.F., Branting, A., Nobile, J.R., Puc, B.P., Light, D., Clark, T.A., Huber, M., Branciforte, J.T., Stoner, I.B., Cawley, S.E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J.A., Namsaraev, E., McKernan, K.J., Williams, A., Roth, G.T. and Bustillo, J. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 475 (7356), 348–52.

Ruberg, F.L. and Berk, J.L. (2012) Transthyretin (TTR) cardiac amyloidosis. *Circulation*. 126 (10), 1286–300.

Saayman, S., Arbuthnot, P. and Weinberg, M.S. (2010) Deriving four functional anti-HIV siRNAs from a single Pol III-generated transcript comprising two adjacent long hairpin RNA precursors. *Nucleic acids research*. 38 (19), 6652–63.

Sagan, S.M., Nasheri, N., Luebbert, C. and Pezacki, J.P. (2010) The Efficacy of siRNAs against Hepatitis C Virus Is Strongly Influenced by Structure and Target Site Accessibility. *Chemistry & Biology*. 17 (5), 515–527.

Saini, H.K., Griffiths-Jones, S. and Enright, A.J. (2007) Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (45), 17719–24.

Sainz, B., Barretto, N., Uprichard, S.L., Bungyoku, Y. and Kitazawa, S. (2009) Hepatitis C Virus Infection in Phenotypically Distinct Huh7 Cell Lines Brett Lindenbach (ed.). *PLoS ONE*. 4 (8), e6561.

Sakurai, Y., Hatakeyama, H., Sato, Y., Hyodo, M., Akita, H. and Harashima, H. (2013) Gene silencing via RNAi and siRNA quantification in tumor tissue using MEND, a liposomal siRNA delivery system. *Molecular therapy : the journal of the American Society of Gene Therapy*. 21 (6), 1195–203.

Salomon, W., Bulock, K., Lapierre, J., Pavco, P., Woolf, T. and Kamens, J. (2010) Modified dsRNAs that are not processed by Dicer maintain potency and are incorporated into the RISC. *Nucleic acids research*. 38 (11), 3771–9.

Salomon, W.E., Jolly, S.M., Moore, M.J., Zamore, P.D. and Serebrov, V. (2015) Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. *Cell*. 162 (1), 84–95.

Sawh, A.N. and Duchaine, T.F. (2012) Turning Dicer on its head. *Nature structural & molecular biology*. 19 (4), 365–6.

Schaefer, B.C. (1995) Revolutions in Rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends. *Analytical Biochemistry*. 227 (2), 255–273.

Schirle, N.T., Kinberger, G.A., Murray, H.F., Lima, W.F., Prakash, T.P. and MacRae, I.J. (2016) Structural Analysis of Human Argonaute-2 Bound to a Modified siRNA Guide. *Journal of the American Chemical Society*. 138 (28), 8694–8697.

Schirle, N.T. and MacRae, I.J. (2012) The crystal structure of human Argonaute2.

Science (New York, N.Y.). 336 (6084), 1037–40.

Schlegel, M.K., Foster, D.J., Kel'in, A. V., Zlatev, I., Bisbe, A., Jayaraman, M., Lackey, J.G., Rajeev, K.G., Charissé, K., Harp, J., Pallan, P.S., Maier, M.A., Egli, M. and Manoharan, M. (2017) Chirality Dependent Potency Enhancement and Structural Impact of Glycol Nucleic Acid Modification on siRNA. *Journal of the American Chemical Society*. 139 (25), 8537–8546.

Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*. 3 (6), 1101–1108.

Schultheis, B., Strumberg, D., Santel, A., Vank, C., Gebhardt, F., Keil, O., Lange, C., Giese, K., Kaufmann, J., Khan, M. and Drevs, J. (2014) First-in-human phase I study of the liposomal RNA interference therapeutic Atu027 in patients with advanced solid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 32 (36), 4141–8.

Schwarz, D.S., Tomari, Y., Zamore, P.D., Schwarz, D.., Bennett, R., Cook, H.., Koppetsch, B.., Theurkauf, W.., Zamore, P.. and Plasterk, R.. (2004) The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Current biology : CB*. 14 (9), 787–91.

Schweizer, P., Pokorny, J., Schulze-Lefert, P. and Dudler, R. (2000) Technical advance. Double-stranded RNA interferes with gene function at the single-cell level in cereals. *The Plant journal : for cell and molecular biology*. 24 (6), 895–903.

Sekijima, Y. (2015) Transthyretin (ATTR) amyloidosis: clinical spectrum, molecular pathogenesis and disease-modifying treatments. *Journal of Neurology, Neurosurgery & Psychiatry*. 86 (9), 1036–1043.

Seo, M.Y., Abrignani, S., Houghton, M. and Han, J.H. (2003) Small interfering RNA-mediated inhibition of hepatitis C virus replication in the human hepatoma cell line Huh-7. *Journal of virology*. 77 (1), 810–2.

Seo, S.B., Zeng, X., King, J.L., Larue, B.L., Assidi, M., Al-Qahtani, M.H., Sajantila, A. and Budowle, B. (2015) Underlying Data for Sequencing the Mitochondrial Genome with the Massively Parallel Sequencing Platform Ion Torrent PGM. *BMC Genomics*. 16 (Suppl 1), S4.

Shepard, A.R., Jacobson, N. and Clark, A.F. (2005) Importance of quantitative PCR primer location for short interfering RNA efficacy determination. *Analytical biochemistry*. 344 (2), 287–8.

Shier, M.K., El-Wetidy, M.S., Ali, H.H. and Al-Qattan, M.M. (2016) Hepatitis c virus genotype 4 replication in the hepatocellular carcinoma cell line HepG2/C3A. *Saudi journal of gastroenterology : official journal of the Saudi Gastroenterology Association*. 22 (3), 240–8.

Shimakami, T., Yamane, D., Jangra, R.K., Kempf, B.J., Spaniel, C., Barton, D.J. and Lemon, S.M. (2012) Stabilization of hepatitis C virus RNA by an Ago2-miR-122 complex. *Proceedings of the National Academy of Sciences of the United States of America*. 109 (3), 941–6.

Singh, R.R., Patel, K.P., Routbort, M.J., Aldape, K., Lu, X., Manekia, J., Abraham, R., Reddy, N.G., Barkoh, B.A., Veliyathu, J., Medeiros, L.J. and Luthra, R. (2014) Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. *British journal of cancer*. 111 (10), 2014–23.

Sioud, M. (2007) RNA interference and innate immunity. *Advanced Drug Delivery Reviews*. 59 (2–3), 153–163.

Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H. and Williams, B.R.G. (2003) Activation of the interferon system by short-interfering RNAs. *Nature Cell Biology*. 5 (9), 834–839.

Smith, D.B., Bukh, J., Kuiken, C., Muerhoff, A.S., Rice, C.M., Stapleton, J.T. and Simmonds, P. (2014) Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology*. 59 (1), 318–327.

Snead, N.M., Wu, X., Li, A., Cui, Q., Sakurai, K., Burnett, J.C. and Rossi, J.J. (2013) Molecular basis for improved gene silencing by Dicer substrate interfering RNA compared with other siRNA variants. *Nucleic Acids Research*. 41 (12), 6209–6221.

Soifer, H.S., Sano, M., Sakurai, K., Chomchan, P., Saetrom, P., Sherman, M.A., Collingwood, M.A., Behlke, M.A. and Rossi, J.J. (2008) A role for the Dicer helicase domain in the processing of thermodynamically unstable hairpin RNAs. *Nucleic acids research*. 36 (20), 6511–22.

Song, J.-J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science*. 305 (5689), .

Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V. and

Dalmay, T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*. 3 (1), 4.

Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J., John, M., Kesavan, V., Lavine, G., Pandey, R.K., Racie, T., Rajeev, K.G., Röhl, I., Toudjarska, I., Wang, G., Wuschko, S., Bumcrot, D., Koteliansky, V., Limmer, S., Manoharan, M. and Vornlocher, H.-P. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature*. 432 (7014), 173–8.

Starega-Roslan, J., Krol, J., Koscianska, E., Kozlowski, P., Szlachcic, W.J., Sobczak, K. and Krzyzosiak, W.J. (2011) Structural basis of microRNA length variety. *Nucleic acids research*. 39 (1), 257–68.

Stein, C.A. and Castanotto, D. (2017) FDA-Approved Oligonucleotide Therapies in 2017. *Molecular Therapy*. 25 (5), 1069–1075.

Stein, P., Rozhkov, N. V., Li, F., Cárdenas, F.L., Davydenk, O., Vandivier, L.E., Gregory, B.D., Hannon, G.J. and Schultz, R.M. (2015) Essential Role for Endogenous siRNAs during Meiosis in Mouse Oocytes Paula E. Cohen (ed.). *PLOS Genetics*. 11 (2), e1005013.

Suhr, O.B., Coelho, T., Buades, J., Pouget, J., Conceicao, I., Berk, J., Schmidt, H., Waddington-Cruz, M., Campistol, J.M., Bettencourt, B.R., Vaishnaw, A., Gollob, J. and Adams, D. (2015) Efficacy and safety of patisiran for familial amyloidotic polyneuropathy: a phase II multi-dose study. *Orphanet Journal of Rare Diseases*. 10 (1), 109.

Suhy, D.A., Kao, S.-C., Mao, T., Whiteley, L., Denise, H., Souberbielle, B., Burdick, A.D., Hayes, K., Wright, J.F., Lavender, H., Roelvink, P., Kolykhalov, A., Brady, K., Moschos, S.A., Hauck, B., Zelenaia, O., Zhou, S., Scribner, C., High, K.A., Renison, S.H. and Corbau, R. (2012) Safe, long-term hepatic expression of anti-HCV shRNA in a nonhuman primate model. *Molecular therapy : the journal of the American Society of Gene Therapy*. 20 (9), 1737–49.

Suzuki, M.T. and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and environmental microbiology*. 62 (2), 625–30.

Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M.W. and Kubista, M. (2015) How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*. 39–16.

Swiecicki, P.L., Zhen, D.B., Mauermann, M.L., Kyle, R.A., Zeldenrust, S.R., Grogan, M., Dispenzieri, A. and Gertz, M.A. (2015) Hereditary ATTR amyloidosis: a single-institution experience with 266 patients. *Amyloid*. 22 (2), 123–131.

Tabernero, J., Shapiro, G.I., LoRusso, P.M., Cervantes, A., Schwartz, G.K., Weiss, G.J., Paz-Ares, L., Cho, D.C., Infante, J.R., Alsina, M., Gounder, M.M., Falzone, R., Harrop, J., White, A.C.S., Toudjarska, I., Bumcrot, D., Meyers, R.E., Hinkle, G., Svrzikapa, N., Hutabarat, R.M., Clausen, V.A., Cehelsky, J., Nochur, S. V, Gamba-Vitalo, C., Vaishnaw, A.K., Sah, D.W.Y., Gollob, J.A. and Burris, H.A. (2013) First-in-humans trial of an RNA interference therapeutic targeting VEGF and KSP in cancer patients with liver involvement. *Cancer discovery*. 3 (4), 406–17.

Te, H.S., Randall, G. and Jensen, D.M. (2007) Mechanism of action of ribavirin in the treatment of chronic hepatitis C. *Gastroenterology & hepatology*. 3 (3), 218–25.

Theotokis, P.I., Usher, L., Kortschak, C.K., Schwalbe, E. and Moschos, S.A. (2017) Profiling the Mismatch Tolerance of Argonaute 2 through Deep Sequencing of Sliced Polymorphic Viral RNAs. *Molecular Therapy - Nucleic Acids*. 922–33.

Treangen, T.J. and Salzberg, S.L. (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature reviews. Genetics*. 13 (1), 36–46.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P. and Sharp, P.A. (1999) Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & development*. 13 (24), 3191–7.

Vlassov, A. V., Korba, B., Farrar, K., Mukerjee, S., Seyhan, A.A., Ilves, H., Kaspar, R.L., Leake, D., Kazakov, S.A. and Johnston, B.H. (2007) shRNAs Targeting Hepatitis C: Effects of Sequence and Structural Features, and Comparision with siRNA. *Oligonucleotides*. 17 (2), 223–236.

Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Kräusslich, H.-G., Mizokami, M., Bartenschlager, R. and Liang, T.J. (2005) Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Medicine*. 11 (7), 791–796.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T. and Patel, D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes.

Nature. 461 (7265), 754–61.

Wang, Y., Luo, J., Zhang, H. and Lu, J. (2016) microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes. *Molecular biology and evolution*. 33 (9), 2232–47.

Wose Kinge, C.N., Espiritu, C., Prabdial-Sing, N., Sithebe, N.P., Saeed, M. and Rice, C.M. (2014) Hepatitis C virus genotype 5a subgenomic replicons for evaluation of direct-acting antiviral agents. *Antimicrobial agents and chemotherapy*. 58 (9), 5386–94.

Yeung, M.L., Yasunaga, J., Bennasser, Y., Dusetti, N., Harris, D., Ahmad, N., Matsuoka, M. and Jeang, K.-T. (2008) Roles for microRNAs, miR-93 and miR-130b, and tumor protein 53-induced nuclear protein 1 tumor suppressor in cell growth dysregulation by human T-cell lymphotrophic virus 1. *Cancer research*. 68 (21), 8976–85.

Yoshikawa, M., Iki, T., Tsutsui, Y., Miyashita, K., Poethig, R.S., Habu, Y. and Ishikawa, M. (2013) 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proceedings of the National Academy of Sciences of the United States of America*. 110 (10), 4117–22.

Yu, M., Peng, B., Chan, K., Gong, R., Yang, H., Delaney, W. and Cheng, G. (2014) Robust and Persistent Replication of the Genotype 6a Hepatitis C Virus Replicon in Cell Culture. *Antimicrobial Agents and Chemotherapy*. 58 (5), 2638–2646.

Zeng, Y. and Cullen, B.R. (2003) Sequence requirements for micro RNA processing and function in human cells. *RNA (New York, N.Y.)*. 9 (1), 112–23.

Zimmermann, T.S., Lee, A.C.H., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M., Judge, A.D., Lam, K., McClintock, K., Nechev, L. V., Palmer, L.R., Racie, T., Röhl, I., Seiffert, S., Shanmugam, S., Sood, V., Soutschek, J., Toudjarska, I., Wheat, A.J., Yaworski, E., Zedalis, W., Koteliansky, V., Manoharan, M., Vornlocher, H.-P. and MacLachlan, I. (2006) RNAi-mediated gene silencing in non-human primates. *Nature*. 441 (7089), 111–114.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*. 31 (13), 3406–15.