Design, Construction, and Screening of an shRNA Library Targeting Human Circular RNAs

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Biochemistry
University of Saskatchewan
Saskatoon

By
Md. Fahmid Islam

Permission to Use Statement

By presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes, may be granted by the professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written consent. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of the materials in this thesis in whole or in part should be addressed to:

Head of the Department of Biochemistry
107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5, Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building
110 Science Place
Saskatoon, Saskatchewan, S7N 5C9, Canada

Abstract

Recent advances in next-generation sequencing (NGS) methods and computational analyses have identified a large class of non-polyadenylated RNA molecules. Further, analysis of these RNAs revealed several unexpected junctions, which do not map to mRNAs, leading to the discovery of circular RNAs (circRNAs). Unlike linear RNAs, circRNAs have no ends and are not sensitive to exoribonucleases, endowing circRNAs with a longer half-life. This enhanced stability has prompted the study of circRNAs as cancer biomarkers. While expression studies are becoming widely used to profile circRNAs in multiple cancers, there are no genome-wide tools available to decrease their levels in cells. Methods that investigate the role of circRNAs through measuring their expression are prone to artifacts as bypassing transcription termination results in RNA concatamers. To better characterize the function of circRNAs, we developed a novel pooled library of ~ 15,000 shRNAs targeting ~ 5,000 circRNAs. We performed a loss-of-function screen with the circRNA shRNA library in a colorectal cancer cell line to systematically identify circRNAs that were required for cell proliferation and survival. During the construction and validation of this library, we also developed a method to improve NGS quality by reducing sequencing failure due to shRNA hairpin and/or heteroduplex formation. Using this approach, we identified and validated several circRNAs essential for the survival of colorectal cancer cell lines. We believe that these essential circRNAs will provide an opportunity to understand cancer biology in a more detailed way, and to design effective cancer therapeutics and diagnostics.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors, Dr. C. Ronald Geyer and Dr. Franco J. Vizeacoumar, for their continuous support and guidance for my research and study. Their immense knowledge and affection for science constantly kept me motivated to find out my answers and solutions that eventually helped me develop both my rational understanding and scientific skills. I whole-heartedly appreciate them for investing many hours to teach me various advanced fields of cancer research, as well as to correct my writings and presentations.

Besides my supervisors, my earnest thanks go to my committee members, Dr. Jeremy Lee, Dr. Scot Leary and Dr. Tony Kusalik for their insightful comments and questions that encouraged me to widen my research perspective. I also wish to thank the Department of Biochemistry for offering some advanced courses and providing me with an opportunity to work as a teaching assistant. These ultimately assisted me to strengthen the background knowledge for my research and perceive the future implications of my work. I would like to thank all the members of the Geyer and Vizeacoumar labs for their dedicated support and thoughtful response in terms of many of my research questions and problems. In particular, I like to convey my regard to Dr. Atsushi Watanabe for training me to the capability where I can carry out the project independently. Moreover, I am thankful to Dr Frederick J. Vizeacoumar and Conor Lazrou for their assistance in the computational part of the research. I will always remember Dr. Sreejit Parameswaran and Wayne Hill for all those stimulating discussions and day-long experiments.

Finally, I am immensely grateful to my parents, Ferdousi and Delwar, as well as my siblings and in-laws, for their patience and inspiration throughout this time of research and study. Last but not the least, my profound thankfulness goes to my wife, Sharmin, for all of her care and encouragement to boost my spirit for overcoming the challenging situations I faced during this time.

Table of Contents

List of Figures

Pages

List of Tables

Pages

List of Abbreviations

15K	15,000
5K	5,000
BSA	Bovine serum albumin
CDR1as	Cerebellar degeneration related protein 1 antisense
CircRNA	Circular RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DMEM	Dulbecco's Modified Eagle Medium
FBS	Fetal bovine serum
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GeCKO	Genome-scale CRISPR Knock Out
GFP	Green fluorescent protein
IDT	Integrated DNA Technologies
miRNA	MicroRNA
mRNA	Messenger RNA
NGS	Next-generation sequencing
PMM1	Phosphomannomutase 1
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RFP	Red fluorescent protein
RPMI	Roswell Park Memorial Institute medium
RPS5	Ribosomal protein S5
SCRIB	Scribbled Planar Cell Polarity Protein
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
SOC	Super optimal broth
WBSCR22	Williams-Beuren syndrome chromosome region 22
YT	Yeast extract tryptone

1.0 Introduction

1.1 Circular RNA: History, Biogenesis and Key Features

Circular RNAs (circRNAs) were first discovered in late seventies (Hsu and Coca-Prados, 1979) and described further by Vogelstein and colleagues in the early nineties (Nigro *et al.*, 1991). For years, circRNAs were thought to originate from low frequency splicing errors. With the advent of next-generation sequencing (NGS) technologies, circRNAs have been confirmed as a new class of RNA that is present across the eukaryotic tree of life (Salzman *et al.*, 2012; Jeck *et al.*, 2013; Salzman *et al.*, 2013; Memczak *et al.*, 2013; Zhang *et al.*, 2013; Wang *et al.*, 2014a; Rybak-Wolf *et al.*, 2015). Until recently, almost all RNA-sequencing methods analyzed the polyadenylated fraction of the transcriptome. These methods use poly (A)-tail purification and are preferred as they eliminate ribosomal RNAs (rRNAs), which constitutes more than 90% of the total RNA in eukaryotic cells. Recent rRNA-depleting NGS methods that use random primer-based cDNA synthesis from the transcriptome, identified a large number of RNAs that are not polyadenylated (Hansen, 2016). These methods revealed unexpected RNA junctions that do not map to the 'linear' genome and represent circRNAs (Salzman *et al.*, 2012). Today, the widespread existence of circRNAs is well established with several biochemical methods and computational algorithms that have been developed to identify them (Wang *et al.*, 2010; Salzman *et al.*, 2012; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Zhang *et al.*, 2014; Rybak-Wolf *et al.*, 2015; Szabo *et al.*, 2015; Hansen *et al.*, 2016). Ongoing efforts are focused on understanding their biogenesis and functions, and more specifically their roles in development and diseases.

While, it is not still clear how most of the circRNAs are formed and how they functions they have in cells, it has been proposed that they are produced by exon scrambling or shuffling, through a non-canonical splicing of pre-mRNAs (also referred to as 'back splicing') that covalently links the 3′- and 5′-ends (Fig. 1.1) (Salzman *et al.*, 2012; Hentze and Preiss, 2013; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Wilusz and Sharp, 2013; Lasda and Parker, 2014; Rybak-Wolf *et al.*, 2015). Recent studies demonstrated that circRNA biogenesis competes with pre-mRNA splicing and that cyclization is dependent on flanking sequences with inverted *Alu* repeats (Jeck *et al*., 2013; Ashwal-Fluss *et al.*, 2014; Liang and Wilusz, 2014; Zhang *et al.*, 2014). *Alu* elements belong to the primate-specific, short-interspersed, nuclear elements (SINEs) that are a family of retrotransposons roughly 0.3 kb in

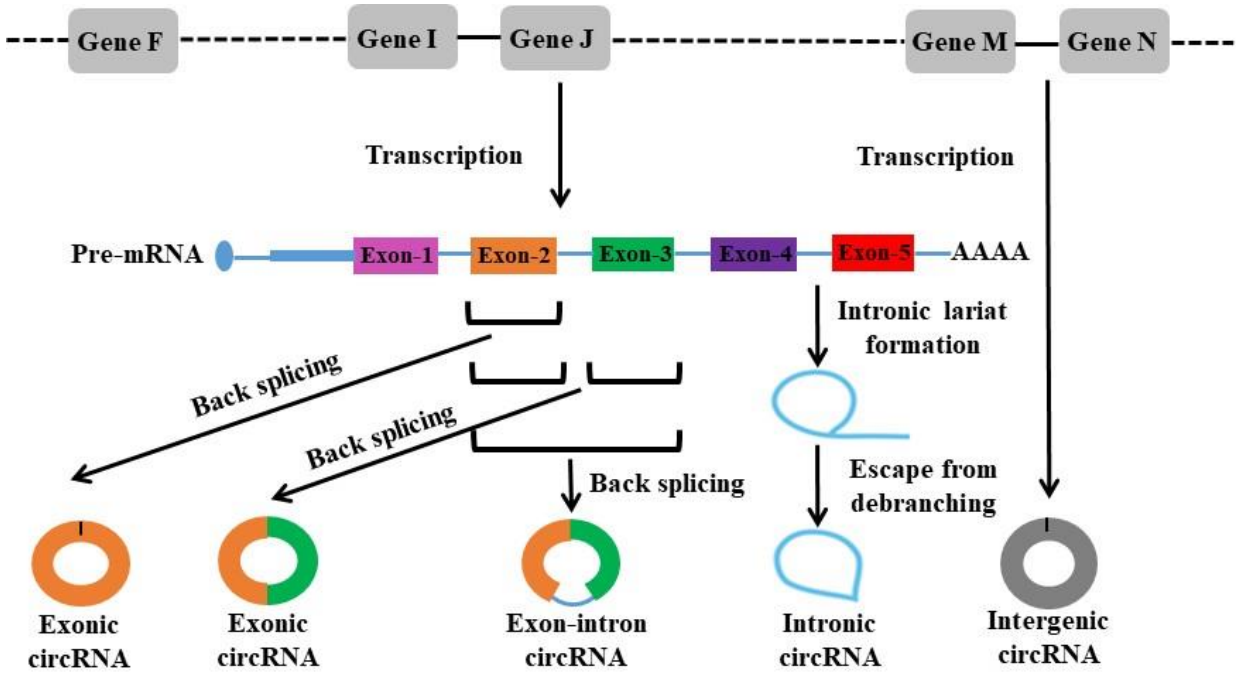

Figure 1.1 Different circRNA biogenesis pathways. Illustration showing different types of circRNA biogenesis pathways. Exonic circRNAs are formed from single or multiple exons by 'back splicing' process during RNA splicing. Intronic circRNAs are formed by intronic lariats that escape from the debranching process during RNA splicing. Exon-intron circRNAs can be formed by any intron trapped inside some exons during the 'back splicing' process. Some intergenic circRNAs are also found that do not have any alignment to known genes.

size and represent 10.5% of the human genome. Besides *Alu* sites, reverse complementary sequences between introns that bracket circRNAs have also been shown to promote cyclization (Ivanov *et al.*, 2015). Furthermore, RNA-binding proteins and spliceosome-dependent cyclization have been proposed to play a role in circRNA biogenesis (Conn *et al.*, 2015; You *et al.*, 2015). In addition to these exonic circRNAs, introns and combinations of exons and introns can produce circRNAs (Salzman *et al.*, 2012; Hentze and Preiss, 2013; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Wilusz and Sharp, 2013; Zhang *et al.*, 2013; Lasda and Parker, 2014; Vicens and Westhof, 2014; Li *et al.*, 2015a; Rybak-Wolf *et al.*, 2015). Based on their biogenesis, circRNAs can be broadly classified as exonic (E-circRNA) or intronic (I-circRNA) or a combination of both exonic and intronic circRNAs (EI-circRNAs) (Fig. 1.1). E-circRNAs are the most abundant class of circRNAs (Salzman *et al.*, 2012; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Rybak-Wolf *et al.*, 2015). It was observed that

circRNAs coming from single or multiple exons (1-5 exons) constitute about 95% of total circRNAs (Memczak *et al.*, 2013). They also found that 85% of circRNAs originate from the sense RNA of protein-coding genes and 10% align to antisense strand of the genes. The remaining 5% circRNAs observed by Memczak *et al.*, 2013 were either I-circRNAs or intergenic circRNAs (Fig. 1.1). Most recently, fusion circRNAs (f-circRNAs) have been identified that are derived from regions of chromosomal translocations (Guarnerio *et al.*, 2016).

Currently, several circRNA databases and online tools are available that archive and characterize circRNA sequences and other information identified in different species (Table 1.1). Mammalian cells contain a large number of circRNAs with enhanced stability and abundance, which are sometimes present at higher levels than their corresponding linear mRNAs (Salzman *et al.*, 2012; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Rybak-Wolf *et al.*, 2015). Jeck *et al.*, 2013 estimated that circRNAs are derived from 15% of actively transcribed genes. Moreover, they show high conservation among mammals, indicating that they have evolutionary and functional importance (Jeck *et al.*, 2013). The size

Table 1.1 Examples of databases and online tools for circRNAs

Database	**Developers**	**Features**
circBase	Glažar *et al.*, 2014	Repository of data sets (genomic annotations and sequences) of circRNAs from different genomic studies
Circ2Traits	Ghosal *et al.*, 2013	Prediction of interaction network among genes of miRNA (microRNA), mRNA (messenger RNA), long non-coding RNA (lincRNA) and circRNA; mapping of disease associated single nucleotide polymorphisms (SNPs) on circRNA loci; identification of Argonaute (AGO) interaction sites on circRNAs
circRNADb	Chen *et al.*, 2016a	Prediction of protein coding potentiality of circRNAs; information about circRNA parental genes in terms of expression and diseases
CircInteractome	Dudekula *et al.*, 2016	Detection of potential circRNAs sponging RNA binding proteins (RBPs); junction-specific primer design for circRNAs; design of siRNAs for silencing of circRNAs; detection of potential internal ribosomal entry sites (IRES) of circRNAs
circNet	Liu *et al.*, 2016	Building miRNA-circRNA-gene network; expression profiling of circRNA isoforms
CSCD	Xia *et al.*, 2017	Exploration of the cancer-specific circRNAs

of predicted and validated circRNAs can range from 100 bases to greater than 10,000 bases with the most species ranging between 100-1500 bases (Salzman *et al.*, 2012; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Salzman *et al.*, 2013; Rybak-Wolf *et al.*, 2015).

1.2 Functions of Circular RNAs

Functions of circRNAs are currently not well understood (Hentze and Preiss, 2013). CircRNAs have been shown to play an important role in modulating gene expression by binding and inactivating miRNAs through a process called sponging, as outlined in Fig. 1.2 (Hansen, 2013). It is evident that miRNAs have key roles in posttranscriptional regulation of gene expression and their sponging influences the overall gene regulatory events of a cell (Bartel, 2014). The miRNA sponging mechanism is based on the presence of multiple binding sites for the same miRNAs on a circRNA. For instance, CDR1as/CiRS (antisense to human *CDR1* gene locus) has more than 60 binding sites for miR-7 and this sponging event is involved in zebrafish brain development (Hansen *et al.*, 2013; Memczak *et al.*, 2013). In another example, the circRNA from the *SRY* gene (antisense to SRY gene locus) plays a role in murine testes development by the sponging of miR-138 (Capel *et al.*, 1993; Hansen *et al.*, 2013).

The importance of miRNA sponging as the main function of circRNAs in gene regulation has been called into question, as many circRNAs do not have sufficient miRNA target sites to support this mechanism of action (Zhang *et al.*, 2013; Guo *et al.*, 2014; Li *et al.*, 2015a). Alternative mechanisms by which circRNAs regulate parental genes in the nucleus and cytoplasm have therefore been proposed (Qu *et al.*, 2015a). It has been observed that some I-circRNAs (e.g. ci-ankrd52) and EI-circRNAs (e.g circEIF3J) are very abundant in the nucleus and show cis-regulatory role on their parental genes (Zhang *et al.*, 2013; Li *et al.*, 2015a). They act as positive regulator of RNA polymerase II and promote transcription of their parental genes. Moreover, some E-circRNAs and mRNAs from their parental genes have common miRNA binding sites, which control the gene regulation in cytoplasm (Li *et al.*, 2015b). Having the same miRNA binding sites creates competition between the circRNA and mRNA that ultimately modulates the gene expression. CircRNAs have also been shown to regulate alternative splicing of pre-mRNAs after transcription (Ashwal-Fluss et al., 2014). It was observed that the abundance of some circRNAs (e.g. circ7780) are correlated with cell proliferation (Bachmayr-Heyda et al., 2015). Another study showed that circHIPK3 regulates

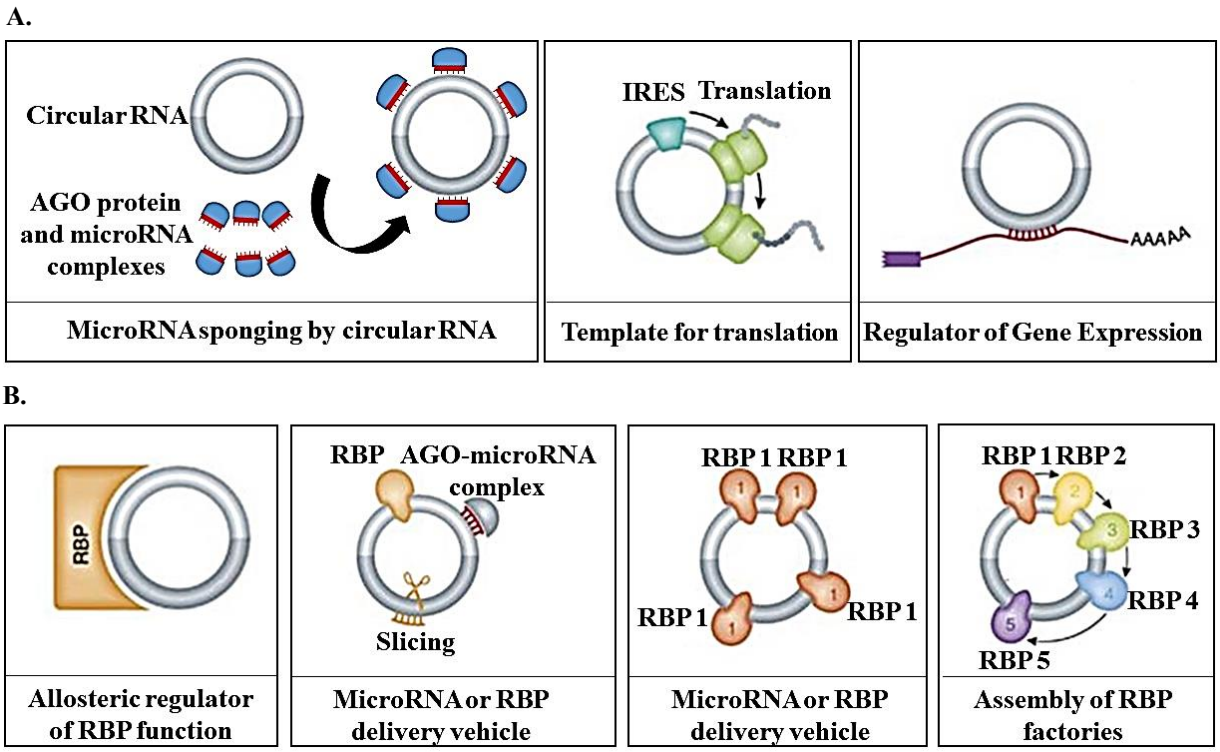

Figure 1.2 Types of circRNA functions. (A) Illustration showing demonstrated functions of circRNAs. CircRNAs can act as miRNA sponges in association with related AGO proteins and thus regulate gene expression. CircRNAs that have an IRES can be translated into proteins. CircRNAs can also interact directly with other mRNAs and regulate their expression. **(B)** Illustration showing other plausible functions of circRNAs. CircRNAs can regulate RBPs, work as delivery vehicles of RBPs and miRNAs, and sponge or assemble RBPs. Adapted from Hentze and Preiss, 2013.

cell growth by sponging multiple miRNAs (Zheng et al. 2016). Overall, these studies indicate that circRNAs play important roles from regulation of gene expression to transcriptional control and cell cycle progression.

Other functions for circRNAs have been proposed, including RBP delivery, RBP sponging, assembly of RBP factories, allosteric regulation of RBP functions, and templating for translation (Fig. 1.2) (Castello *et al.*, 2012; Hentze and Preiss, 2013; Wilusz and Sharp, 2013, Ashwal-Fluss *et al.*, 2014; Jeck *et al.*, 2014; Lasda and Parker, 2014; Salzmn, 2014). Though circRNAs were initially thought to be non-coding RNA (ncRNA), studies have shown that they can be translated *in vitro* and *in vivo*, indicating an unexplored layer of gene activity (Chen and Sarnow, 1995; Pamudurti *et al.*, 2017).

Despite these studies described circRNA functions, it is still not clear how the network of multiple interactions involving circRNA-miRNA-mRNA axes is maintained and controlled throughout circRNA processing and beyond. Many questions about the function of circRNAs remain, from their role in diseases like cancer to their epigenetic heritability. Understanding the purpose of circRNAs may open doors to therapies for diseases as well as a better understanding of their involvement in biological processes.

1.3 Circular RNAs and Human Cancer

Proper understanding of circRNAs and their functions will lead to better decoding of complex diseases (Chen *et al.*, 2016b). Specifically, circRNAs are a new avenue of investigation to understand cancer (Li *et al.,* 2015c). It is already evident that miRNAs play pivotal roles in different stages of cancer through regulating the expression of tumor suppressor genes and oncogenes (Lee and Dutta, 2009). In some malignancies, miRNA expression is often deregulated. The sponging effect of circRNA on miRNAs indicates a direct involvement in this regard (Hansen *et al.*, 2013). For example, cir-ITCH circRNA derived from the *ITCH* gene presents a sequence enriched with three miRNA-binding sites (miR-7, miR-17 and miR-214) and interactions of cir-ITCH with miR-7, miR-17, and miR-214 might increase the level of ITCH, which inhibits the Wnt/β-catenin pathway (Li *et al.*, 2015b). The expression of cir-ITCH in esophageal squamous cell carcinoma is very low, indicating that tumors downregulate cir-ITCH to enhance proliferation (Li *et al.*, 2015b). It also acts in colorectal cancer by targeting Wnt/β-catenin pathway (Huang *et al.*, 2015). In contrast, circTCF25 is upregulated in bladder cancer and imparts its effect through circTCF25-miR-103a-3p/miR-107-CDK6 pathway (Zhong *et al.*, 2016). Recently, Burton Yang's group demonstrated that circRNAs can bind to proteins involved in controlling cell cycle, senescence, and tumor progression (Du *et al.*, 2016; Yang *et al.*, 2016; Du *et al.*, 2017). In addition, they showed that circRNAs can promote nuclear translocation of oncogenic proteins (Yang *et al.*, 2017a; Yang *et al.*, 2017b). Some f-circRNAs play roles in acute promyelocytic leukemia through regulating PI3K and MAPK signal transduction pathways (Guarnerio *et al.*, 2016).

Unlike linear RNA, 3′- and 5′-ends of circRNA are not exposed, making them less sensitive to ribonucleases and imparting a longer half-life. This enhanced stability has led to the use of circRNAs as clinical biomarkers, especially in cancer (Memczak, 2015; Zhang *et al.*,

2017). Recently, several studies have found differential expression of circRNAs between normal and cancer tissues (Table 1.2) (Qu *et al.*, 2015b; Ahmed *et al.*, 2016; Song *et al.*, 2016). Interestingly, enrichment of certain circRNAs in exosomes can potentially benefit cancer diagnosis as a biomarker (Li *et al.*, 2015e). One of the more studied circRNAs, CDR1as, is described by Ajay Goel's group as a promising prognostic biomarker and a potential therapeutic target in colorectal cancer due to its ability to act as a sponge for miR-7, a known tumor suppressor (Weng *et al.*, 2017).

1.4 Functional Genomics, Pooled Screening, and Gene Essentiality

Recent advancements in sequencing technologies and their applications in functional genomics have significantly broadened our understanding of cellular functions and our ability to perform translational science (Islam *et al.*, 2017). These technologies often involve sequencing a pool of DNA oligonucleotides that are unique in nature (Islam *et al.*, 2017). For example, large-scale, genome-wide screens using pooled shRNA or CRISPR libraries and subsequent NGS identified the unique shRNA or sgRNA sequences that affect cell viability (Fig. 1.3) (Blakely *et al.*, 2011; Ketela *et al.* 2011; Marcotte *et al.*, 2012; Bassik *et al.*, 2013; Koike-Yusa *et al.*, 2013; Mali *et al.*, 2013; Vizeacoumar *et al.*, 2013; Shalem *et al.*, 2014; Wang *et al.*, 2014b; Paul *et al.*, 2014; Hart *et al.*, 2015; Paul *et al.*, 2016). These methods are increasingly being applied to detect therapeutically relevant synthetic lethal targets (Luo *et al.*, 2009; Brough *et al.*, 2011; Vizeacoumar *et al.*, 2013; Bajrami *et al.* 2014; Cermelli *et al.*, 2014; Paul *et al.*, 2014; Van Der Meer *et al.*, 2014; Paul *et al.*, 2016) or cancer-specific essential

Table 1.2 Examples of of circRNAs in cancer

Cancer name	CircRNA name	Expression level	Study group
Cutaneous squamous cell carcinoma	hsa_circ_0035381	Up	Sand *et al.*, 2016
	hsa_circ_0022383	Down	
Gastric cancer	hsa_circ_002059	Down	Li. *et al.*, 2015d
Colorectal cancer	hsa_circ_001988	Down	Wang *et al.*, 2015
Hepatocellular carcinoma	hsa_circ_0001649	Down	Qin *et al.*, 2016
Laryngeal squamous cell cancer	hsa_circ_104912	Down	Xuan *et al.*, 2016
	hsa_circ_100855	Up	

genes (Paddison *et al.*, 2004; Luo *et al.*, 2008; Schlabach *et al.*, 2008; Silva *et al.*, 2008; Barbie *et al.*, 2009; Marcotte *et al.*, 2012; Koike-Yusa *et al.*, 2013; Shalem *et al.*, 2014; Wang *et al.*, 2014b; Hart *et al.*, 2015; Munoz *et al.*, 2016). These novel interactions reveal potential targetable vulnerabilities of malignant cells and have resulted in the initiation of several clinical trials (NCT01791309; NCT01750918; NCT01719380). Similarly, NGS technologies are also used in selection techniques such as phage display, mRNA display, yeast display, and aptamer libraries (Ravn *et al.*, 2013; Matochko and Derda, 2015; Van Blarcom *et al.*, 2015; Jalali-Yazdi, 2016; Tolle and Mayer, 2016). A common theme in all of these sequencing reactions is that they depend on mixed-oligonucleotide PCR reactions wherein unique sequences are binned by molecular barcodes distinct to each sequence, allowing multiplexing (Islam *et al.*, 2017).

While these sequencing methods are increasingly used in large core facilities, there are a number of challenges that impede their widespread usage in standard labs where cost-effective

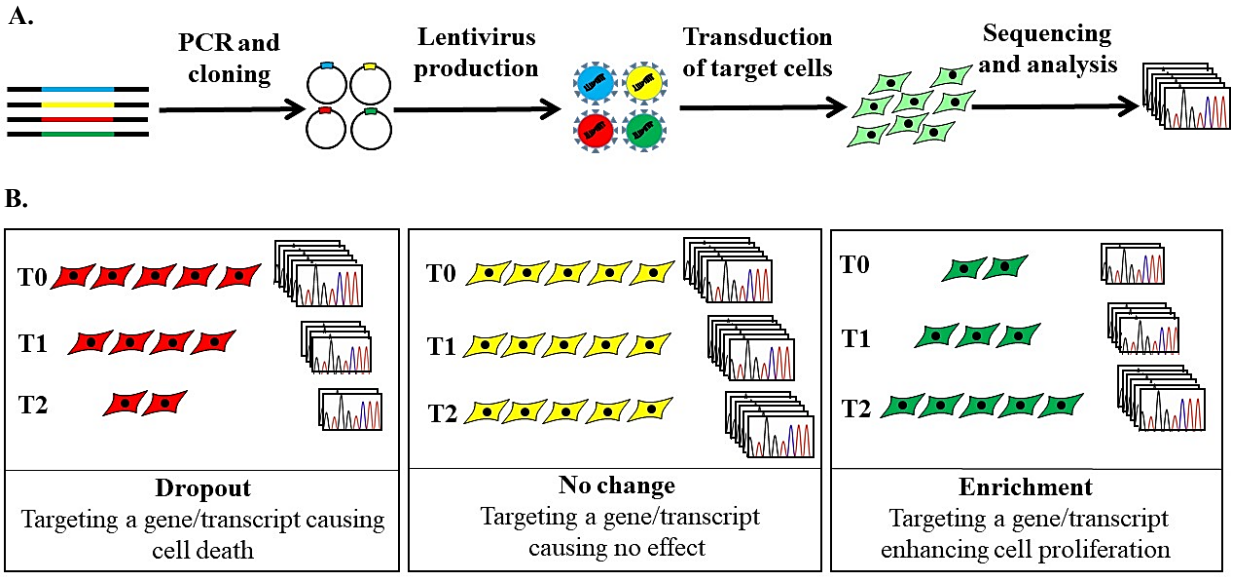

Figure 1.3 Library-based pooled screening pipeline. **(A)** Schematic showing pooled screening methodology with a library (e.g. shRNA or sgRNA library). Specifically designed pooled oligonucleotides (shRNAs or sgRNAs targeting genes or transcripts) are amplified and cloned into an appropriate plasmid. The plasmid library is then packaged into lentivirus to transduce target cells. The library is prepared from genomic DNAs from different time points (T0, T1, and T2) of transduced cell passages for NGS. **(B)** Illustration showing the relation between NGS data and cellular phenotypes. The data are analyzed and corresponding dropout and enriched sequences are identified to determine the effect of target gene/transcript on cellular phenotypes (e.g. cell death).

bench-top sequencers are routinely employed (Islam *et al.*, 2017). One reason for this is that most of these libraries are extremely large and these instruments do not provide the adequate read numbers required for expected sequence coverage (Islam *et al.*, 2017). The availability of sub-libraries that target a small subset of genes (such as ion channels) can alleviate this issue and enhance the feasibility of using low-to-medium throughput sequencers (Islam *et al.*, 2017). However, the formation of secondary structures and mixed heteroduplex templates results in a significant challenge, as these structures reduce the number of useable sequences, in a technology which already experiences limited throughput (Gorbacheva *et al.*, 2015). The development of methods to mitigate sequencing failures will not only enhance the routine application of these techniques in standard labs, but will also increase the throughput and multiplexing capabilities in large core facilities (Islam *et al.*, 2017).

Sequencing failure primarily occurs due to the formation of hairpins and heteroduplexes, as illustrated in Fig. 1.4 (Islam *et al.*, 2017). The formation of heteroduplex is common when sequencing a library of DNA variants derived from the same parental or closely related templates (Islam *et al.*, 2017). Particularly during PCR amplification of mixed-oligonucleotides, annealing of similar types of library sequences results in heteroduplex formation when there is a primer shortage (Ruano and Kidd 1992; Thompson *et al.*, 2002; Meyer and Kircher, 2010; Bowman *et al.*, 2013; Liu *et al.*, 2014; Rentero-Rebollo *et al.*, 2014; Brandariz-Fontes *et al.*, 2015; Gorbacheva *et al.*, 2015). This heteroduplex usually contaminates the intended library and reduces the quality of sequencing due to incomplete, low quality, and polyclonal reads (Islam *et al.*, 2017). Hairpin structures result from palindromic sequences and also lead to similar inadequate reads as heteroduplex formations (Miyagishi, 2004; Kieleczawa, 2005; Kieleczawa, 2006; McIntyre and Fanning, 2006; Gorbacheva *et al.*, 2015). To improve NGS quality, we developed a strategy to reduce sequencing failure due to shRNA hairpin and/or heteroduplex formation as described by Islam *et al.*, 2017. We optimized our methods using a minimally pooled shRNA library consisting of ~15,000 unique shRNA clones. We also performed additional validation assays of our method by applying to a pooled CRISPR library called GeCKO that targets early consecutive exons for genome editing (Shalem *et al.*, 2014).

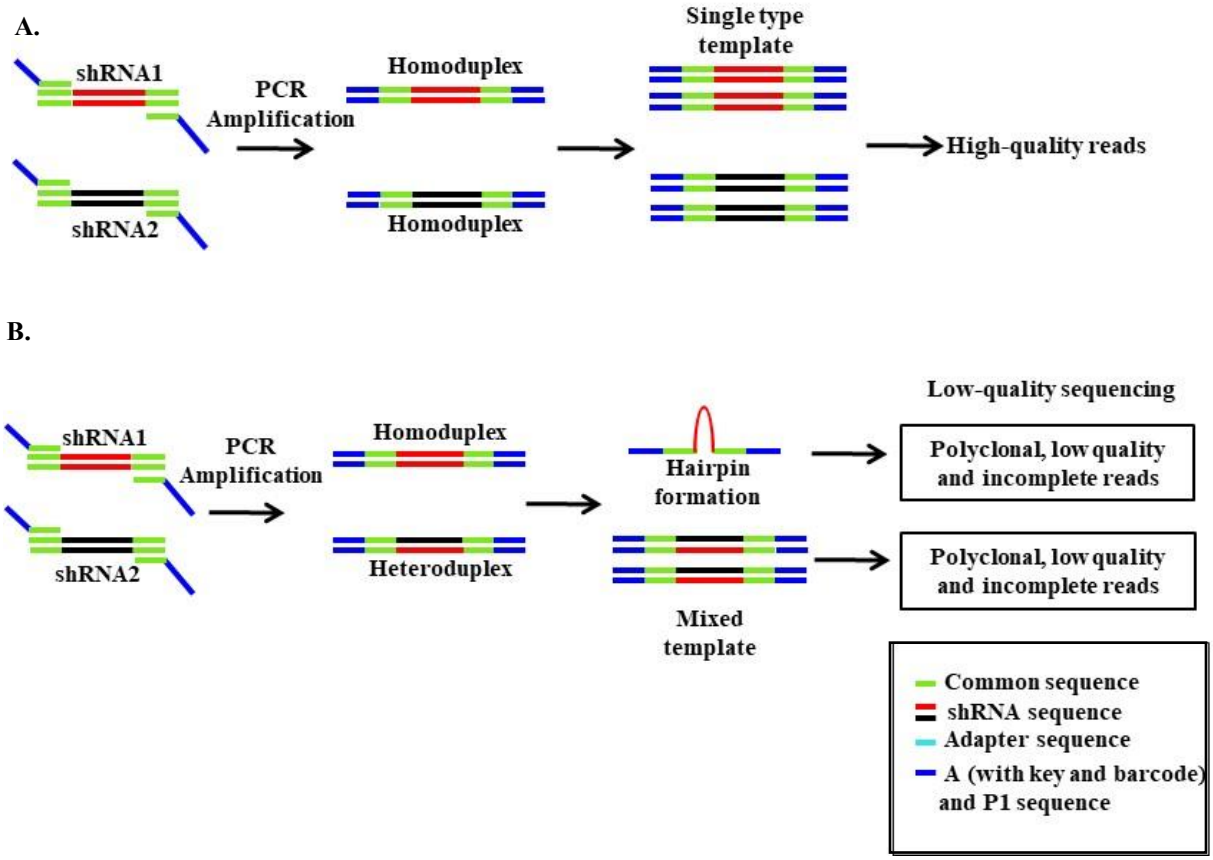

Figure 1.4 Challenges with shRNA library sequencing. (A) Schematic showing expected PCR product when amplifying a mixed-oligo library. **(B)** Schematic showing formation of secondary structure (hairpin structure) and heteroduplex (mixed template due to primer shortage during high number of PCR cycles) resulting in low quality sequencing reads. Adapted from Islam *et al.,* 2017.

1.5 Circular RNA Essentiality in Cancer through Pooled Screening

While the differential expression of circRNAs in cancer-related pathways has been explored, it is not clear if these changes in circRNA expression are a cause or consequence of these malignancies. Gene expression studies in normal and tumor tissues over the past two decades have shown that not all differentially regulated genes are essential for the survival of cancer cells. Therefore, establishing the cancer-specific essentiality of individual circRNAs will provide a novel 'therapeutic dimension' for the role of circRNAs in cancers. In addition, current methods often study the role of circRNAs by using circRNA overexpression plasmids. Mammalian plasmids normally contain the circularized exon(s) along with flanking splicing signals and intronic sequences, which harbor inverted repeats to facilitate their splicing into a

circRNA (Hansen *et al.*, 2013; Ashwal-Fluss *et al.*, 2014; Liang and Wilusz, 2014; Li *et al.*, 2015b). Generally, the transcription termination signals in the plasmid are bypassed, where RNA polymerase will continue to transcribe around the entire plasmid, generating a concatamer of the RNA sequence with undesired transcripts. This technical artifact will contain a scrambled junction and might appear identical to a bona fide circRNA, depending on the approach used for its detection (Barrett *et al.*, 2015; Barrett and Salzman, 2016). This can lead to off-target effects on the cell and spurious circRNA quantification by qRT-PCR. To overcome these constraints, we developed a strategy to rapidly perform genome-wide, loss-of-function screens to systematically identify circRNAs required for cell proliferation and survival.

2.0 Rationale, Hypothesis and Objectives

Currently there are no genome-wide efforts available to disrupt the function of individual circRNAs and understand their role in cancers. To address this, shRNA libraries are needed that specifically target circRNAs.

2.1 Hypothesis

We hypothesized that circRNAs whose loss-of-function cause lethality to cancer cells will provide an opportunity to characterize and understand the biology of a subset of circRNAs in human cancer. To test this hypothesis, we developed a circRNA shRNA library, screened a colorectal cancer cell line for circRNA essentiality with the circRNA shRNA library, and validated some of the top essential circRNAs.

2.2 Objectives

1. To construct an shRNA library that specifically targets human circRNAs
2. To screen a human cancer cell line using this library and identify essential circRNAs
3. To validate the essentiality of the identified circRNAs

3.0 Methods

All reagents and kits were used and stored according to the standard procedures provided by the manufacturers. DNA/RNA samples were contained in DNAase-free ddH_2O and stored at -20˚C for future use. Cell pellets (bacterial and mammalian) from experiments were stored at -80˚C before plasmid, genomic DNA (gDNA) or RNA isolation. Bacterial cell stocks and lentivirus were stored at -80˚C. Mammalian cell stocks were stored at -150˚C. Gentle vortexing and/or pipetting were required at different steps to mix samples and reagents properly. Filtered micropipette tubes or tips were used to reduce any biological or chemical contamination. Different types of containers (e.g. flasks, tubes, plates) and common instruments (e.g. shakers, microwave ovens, centrifuges) were used in the experiments. Standard biosafety guidelines and aseptic practices were followed at every step of the research work.

3.1 List of Uncommon Reagents

Table 3.1 List of uncommon reagents

Reagents	**Supplier Address**
7-Aminoactinomycin D (7-AAD)	BD Biosciences (Mississauga, ON, Canada)
Crystal violet	Sigma-Aldrich (St. Louis, MO, USA)
Polybrene	Sigma-Aldrich (St. Louis, MO, USA)
Polyethyleneimine	Sigma-Aldrich (St. Louis, MO, USA)
Puromycin	Thermo Fisher Scientific (Burlington, ON, Canada)

3.2 Design, Construction and Evaluation of shRNA Library Targeting Circular RNAs

3.2.1 Retrieval of Circular RNA Sequence Information from Database

Information for human circRNAs from different genomic studies was collected from the 'circBase' database (http://www.circbase.org/). In this database, the 'table browser' option was selected to download 'Human (hg19)' circRNAs from four major genomic studies performed prior to 2014 (Memczak *et al.*, 2013; Jeck *et al.*, 2013; Salzman *et al.*, 2013; Zhang *et al.*, 2013). All other parameters were set at default. The separate 'xlsx' files from these studies contained information on the organism, genomic position, strand, circRNA ID, genomic length,

spliced length, samples, scores, repeats, annotation, best transcript, gene symbol, and name of the circRNA study. The corresponding circRNA sequences were retrieved from 'fasta' files of 'Putative spliced circRNA sequences' for 'hg19 assembly' under 'downloads' option in the circBase.

3.2.2 Developing Computational Methods for Designing CircRNA shRNA Library

The circRNA sequences were used to design an shRNA library based on the method described at the 'Broad Institute's TRC Portal. All scripts for the design of the library were written in Python-2.7.9. The design process of the circRNA shRNA library is outlined in Fig. 3.1. Candidate sequences that were 21 bases long (21-mers) for shRNAs were generated from each circRNA sequence (Fig. 3.2A). These candidate sequences were considered as sense strands of the respective shRNAs. The reverse 21-mers of the corresponding candidate sequences were also generated and considered as antisense strands of the shRNAs (Fig. 3.2A). Only the candidate sequences that were highly specific for a given circRNA were retained. For example, the 21-mers that target more than one circRNA were eliminated. Likewise, if the reverse 21-mers of a 21-mer candidate sequence matched to any regions within the same circRNA and/or a different circRNA that 21-mer was eliminated (Fig. 3.2B). Thus, we made sure that none of the shRNAs targeted more than one circRNA. We then calculated an intrinsic score for each remaining 21-mer candidate sequence using a standard 'Rule Set 9' as designed by the Broad Institute (https://portals.broadinstitute.org/gpp/public/resources/rules). 'Rule Set 9' either penalizes or rewards features predicting successful knockdown and clone-design considerations as described in Figure 3.2C. The 'Rule Set 9' criteria to design effective shRNAs are listed in Table 3.2. It is a standard procedure in the design of shRNAs that candidate sequences having a score < 4.96 were eliminated. To reduce off-target effects of the circRNA shRNA library that arise from targeting linear transcripts, sense 21-mer sequences and their antisense 21-mer sequences were matched with all human mRNA sequences downloaded from 'UCSC Genome Browser' (https://genome.ucsc.edu/index.html). Candidate sequences were removed if these 21-mers and/or their reverse 21-mers had 100% identity to the linear mRNA transcripts (Fig. 3.2D). Thus it was ensured that no shRNA in the circRNA shRNA library targeted any human mRNAs. As most shRNAs targeting the body of the circRNA mapped to the exonic regions (and most of the circRNAs were exonic in our list), this resulted

Table 3.2 'Rule Set 9' by Broad Institute for designing shRNAs

Rule	**Description**
aaStart9	Exclude any candidate beginning with AA (score = 0)
fourRow9	Exclude any candidate containing a run of four of the same base in a row (score = 0)
gcScore9	Exclude candidates with extreme GC percentage (GC <= 25% or >60 %); promote candidates with GC between 25-55% (score = 3); if GC > 55% and <= 60% then score = 1 (neutral)
nonGATC9	Exclude any candidate containing ambiguous bases (e.g. N) (score = 0)
restrictionSite9	Exclude any candidate containing certain restriction sites: ...GGTACC..., ...GAATTC..., ...CTCGAG..., ...CATATG..., ...ACTAGT..., ...GGTAC, ...GAATT, GTACC..., TACC..., CTAGT...
sevenGC9	Exclude any candidate with a run of 7 C/G bases (score = 0)
stemLoopStem	Penalize candidates that can form an internal stem-loop (score = 0.1) (minimum stem length = 5, minimum loop size = 4)
threePrimeCalmp6	Give precedence to candidates with weaker base-pairing at positions 15-20 (priority on position 17-19); score = 5 if all 6 positions are A or T, decreasing to 0.1 if all 6 are G/C. Score drops off steeply as the number of A/T bases decreases.

in different numbers of shRNAs per circRNA, targeting the back-spliced region alone. For example, some circRNAs had just one shRNA targeting the back spliced junction while others had eight to ten overlapping shRNAs per circRNA. We decided to have a minimum of one shRNA and a maximum of four shRNA targeting the back spliced region, per circRNA. Based on their intrinsic scores, only the top four sequences were retained for each circRNA if it has more than four shRNAs (Fig. 3.2E). Final oligonucleotide sequences of shRNAs, consisting of 82 base pairs, were designed to contain the hairpin with at 21-mer stem and an *Xho*I site (six base pairs) in the loop between the 21-mers. The hairpin was flanked by common sequences for PCR amplification (17 base pairs on either side). The *Xho*I site allowed restriction digestion to eliminate secondary structure (hairpin and/or cruciform). A common design of the shRNA-oligonucleotide sequence is presented in Fig. 3.2F. Oligonucleotides encoding these sequences were synthesized using DNA microarray technology (LC Sciences). The number of the

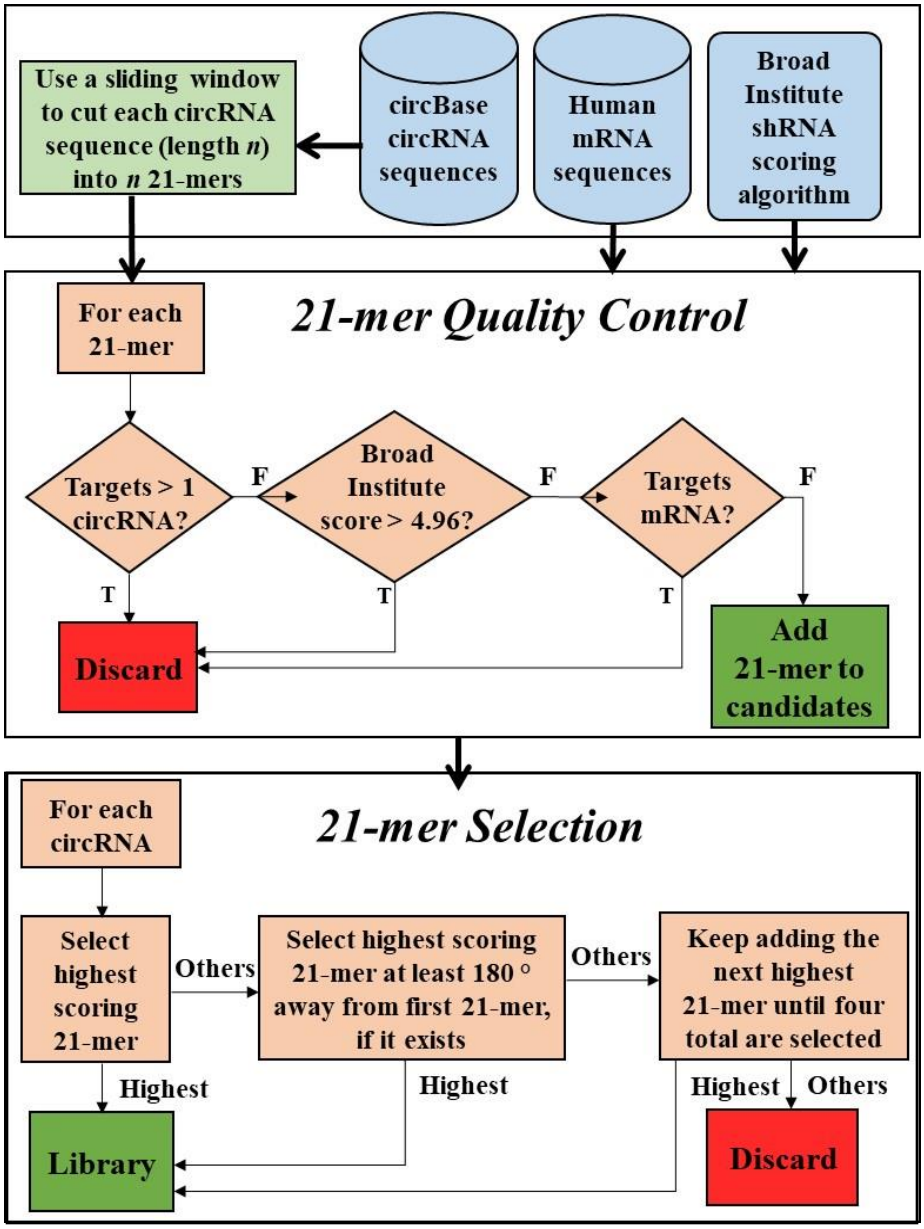

Figure 3.1 Design process of circRNA shRNA library. Flow chart showing the design process of circRNA shRNA library. Python codes were generated for each step of the pipeline to design shRNAs to efficiently target specific circRNAs. Sequences of circRNAs from 'circBase' and sequences of mRNAs from 'USCSC Genome Browser' were used as inputs. 'Rule Set 9' by Broad Institute was used to calculate intrinsic scores of candidate 21-mer sequences. After several filtering and selections steps, the retained 21-mers were used to create the circRNA shRNA library.

oligonucleotides was adjusted to maintain the DNA chip coverage set by the company. Two chips were used to synthesize two sets of the oligonucleotides that were eventually eluted from the chips and pooled together.

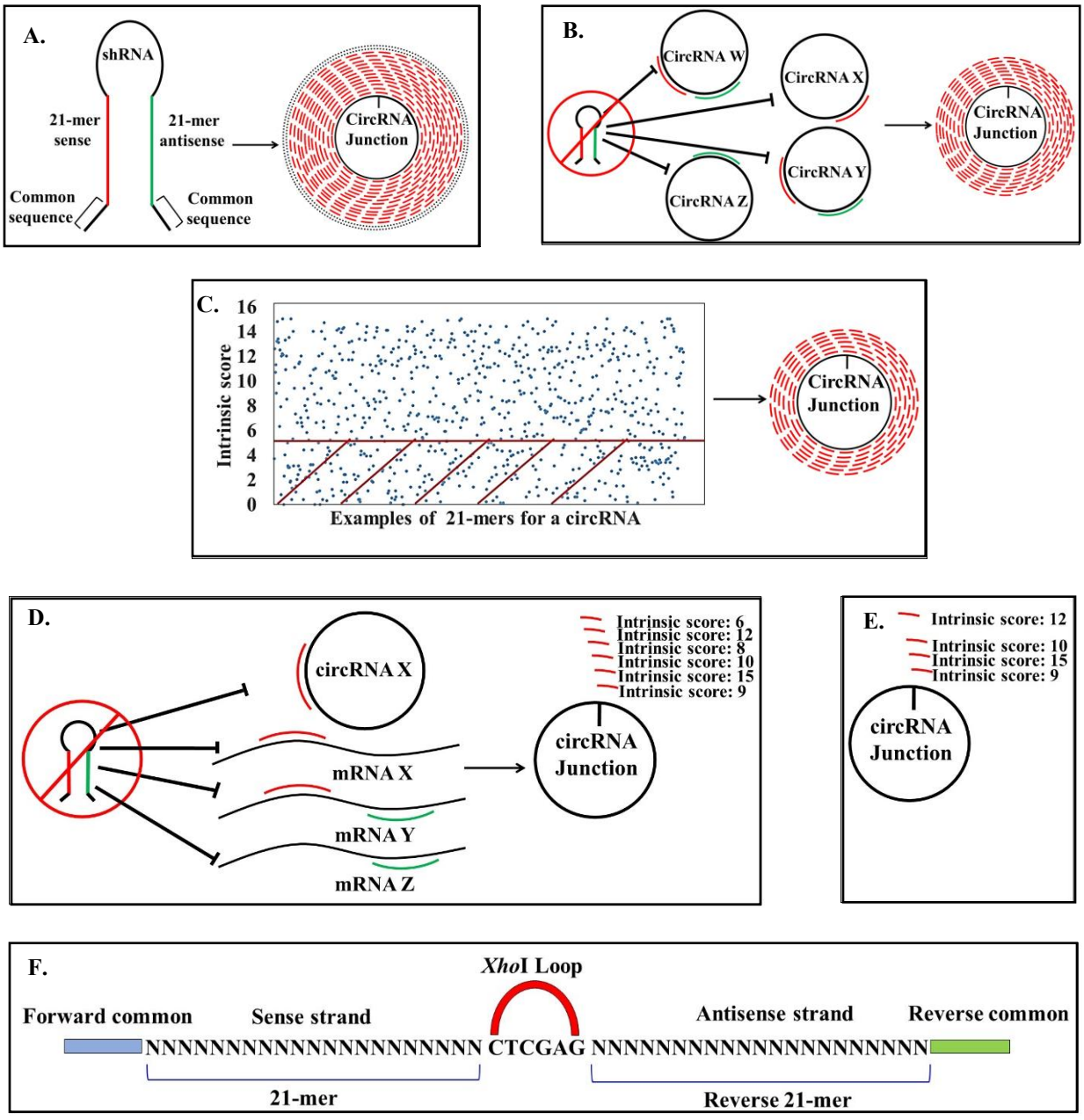

Figure 3.2 Steps in the design process of circRNA shRNA library. (A) Illustration showing generation of all possible 21-mers (sense strand of shRNA) and corresponding reverse 21-mers (antisense strand of shRNA) for a circRNA. **(B)** Illustration of filtering 21-mers that target more than one circRNA. If the corresponding reverse 21-mers target the same and/or any other circRNA, 21-mers were also filtered. **(C)** Bar diagram and illustration showing selection of 21-mers based on intrinsic score calculated with 'Rule Set 9'. **(D)** Illustration of filtering 21-mers that (and/or their reverse 21-mers) target any mRNAs. **(E)** Illustration showing final selection of top four 21-mers based on intrinsic scores. **(F)** Illustration of the design of an shRNA oligonucleotide consisting of 21-mer, reverse 21-mer, *Xho*I loop, and forward and reverse common sequences.

3.2.3 PCR Amplification of CircRNA shRNA Library Oligonucleotides

The synthesized oligonucleotides of the circRNA shRNA library were PCR-amplified with forward and reverse primers that have Gibson Assembly linkers (Table 3.3). Conditions were optimized to perform a PCR amplification using 4.1 ng/µL template oligonucleotide mix (9.1×10^9 copies of oligonucleotides in 1 µL) dissolved in 25 µL of DNAase-free ddH_2O. For the PCR, 10 µL of 5 X HF Buffer (New England BioLabs), 2 µL of 10 mM dNTPs (Thermo Fisher Scientific), 5 µL of 10 µM Forward Primer, 5 µL of 10 µM Reverse Primer and 1 µL of Phusion High-Fidelity DNA Polymerase (New England BioLabs) were mixed with 10 µL of 10 X diluted template oligonucleotide mix (> 6×10^4 times coverage of the library). DNAase-free ddH_2O was added to bring the final volume to 100 µL. The 100 µL PCR was divided into two reactions and PCR was performed with a Thermal Cycler (Applied Biosystems). The temperature profile for the PCR was set as 1 min at 98˚C and 30 cycles of amplification (1 min at 98˚C, 15 sec at 60˚C, 1 min at 72˚C) and reactions were pooled together afterwards. An additional reaction mix of 100 µL was prepared using 20 µL of 5 X HF Buffer, 2 µL of 10 mM dNTPs, 20 µL of 10 µM Forward Primer, 20 µL of 10 µM Reverse Primer, 1 µL of Phusion High-Fidelity DNA Polymerase and 37 µL of DNAase-free ddH_2O. Three reactions were prepared by mixing 25 µL of the first PCR reaction and 25 µL of additional reaction mix (25 µL of first PCR reaction was kept for checking the size). An extra PCR cycle was run as 2 min at 98ºC, 20 sec at 60ºC, and 2 min at 72ºC. Reactions were pooled together (25 µL of first the PCR reaction was kept for checking the amplicon size). First and second PCR reactions (25 µL) were mixed with 5 µL of 6 X DNA Gel Loading Dye (Thermo Fisher Scientific) and resolved on a 2% UltraPure™ Agarose gel (Thermo Fisher Scientific) at 80 V for 50 minutes in a Mini-Sub® Cell GT Systems (Bio-Rad). The agarose gel was stained with SYBR® Safe

Table 3.3 List of Primers for circRNA shRNA library construction (IDT)

ID	Oligonucleotide	Sequence (5′-3′)
1.	Gibson Assembly linker F.P. (*Age*I)	CTTTATATATCTTGTGGAAAGGACGAAACA[3]*CCTCACCA CTCTCCACC*
2.	Gibson Assembly linker R.P. (*EcoR*I)	ATGAATACTGCCATTTGTCTCGAGGTCGAG[4]GGTGGTGT GGTGTAAGG

F.P.: Forward primer, R.P.: Reverse primer, *Common sequence 1*, Common sequence 2, *Age*I linker[3], *EcoR*I linker[4]

DNA Gel Stain (Thermo Fisher Scientific) in 1 X TAE (Thermo Fisher Scientific). The 142 bp expected amplicon was detected on Gel Doc™ XR+ Imager (Bio-Rad) using 100 bp DNA Ladder (Thermo Fisher Scientific). From the second PCR reaction, 125 µL product was mixed with 25 µL of 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel at 80 V for 50 minutes in Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 142 bp expected band was excised on a UV Transilluminator (Thermo Fisher Scientific). The expected product was purified by QIAquick Gel Extraction Kit (QIAGEN), according to the manufacturer's protocol and eluted in 50 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using NanoDrop™ 2000/2000c (Eppendorf).

3.2.4 Cloning of CircRNA shRNA Library Oligonucleotides into an shRNA Expression Plasmid

The circRNA shRNA library amplicon was cloned into pLKO.1-TRC plasmid (8,901 bp, Addgene). pLKO.1-TRC plasmid was digested with *Age*I and *EcoR*I to release the ~ 1.9 kb stuffer fragment (Fig. 3.3). To perform this reaction, 5 µg of pLKO.1 was mixed with 10 µL of 10 X FastDigest Buffer (Thermo Fisher Scientific), 5 µL of FastDigest *Age*I (Thermo Fisher Scientific), 5 µL of FastDigest *EcoR*I (Thermo Fisher Scientific), and DNAase-free ddH_2O was added to bring the volume to 100 µL. The reaction was incubated at 30˚C for 30 minutes on an Isotemp™ Incubator (Thermo Fisher Scientific). The digested pLKO.1-TRC plasmid was mixed with 20 µL of 6 X DNA Gel Loading Dye and was resolved on an 1% UltraPure™ Agarose gel at 90 V for 1 hour in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The ~ 7 kb expected band was detected on Gel Doc™ XR+ Imager using 1 kb DNA Ladder (Thermo Fisher Scientific), cut out on a UV Transilluminator, purified by QIAquick Gel Extraction Kit according to the manufacturer's protocol, and eluted in 50 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using a NanoDrop™. To perform the cloning, 100 ng of *Age*I- and *EcoR*I- digested, purified pLKO.1 (~ 7 kb, 1.3 x 10^{10} copies) and 10 ng of PCR-amplified library oligonucleotides (142 bp, 6.4 x 10^{10} copies) were mixed (plasmid: insert molar ratio of 1: 5 in 10 µL). Three reactions were prepared by mixing this 10 µL of digested-plasmid/insert mixture and 10 µL of 2 X NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs) and

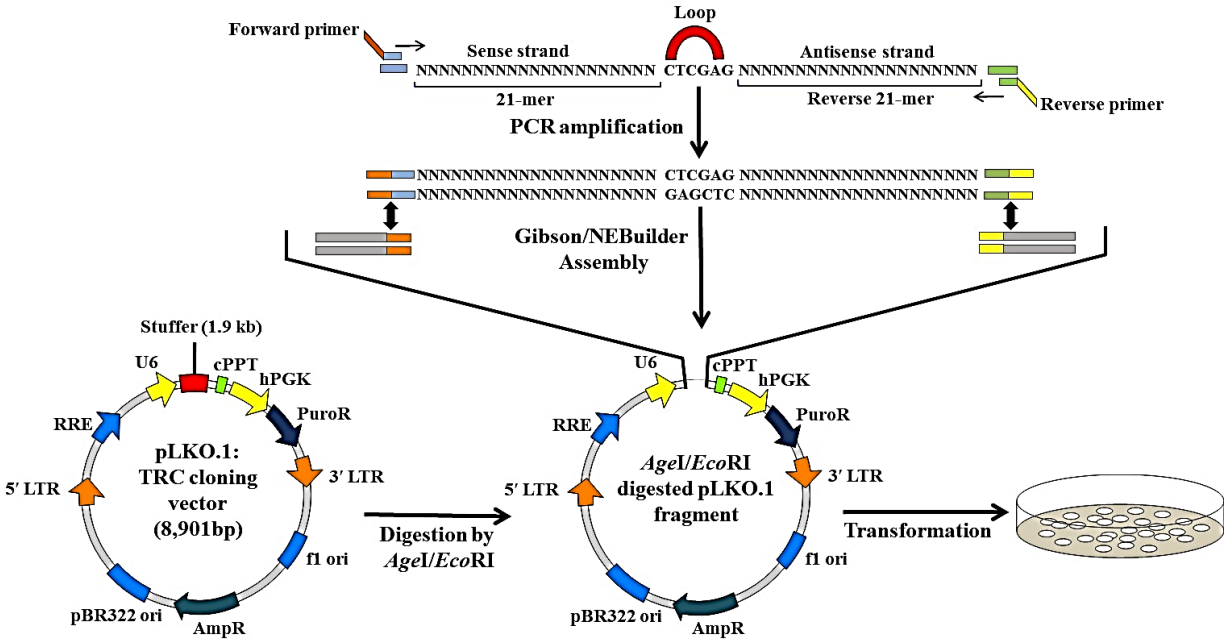

Figure 3.3 Construction of circRNA shRNA library. Illustration showing the construction method of circRNA shRNA library. Template oligonucleotide mix of shRNA library was PCR-amplified with primers having Gibson Assembly linkers. The stuffer fragment was released from pLKO.1 plasmid by double-digestion with *Eco*RI and *Age*I. Library amplicon was cloned into the plasmid using Gibson/NEBuilder Assembly. The cloned plasmid was used to transform ultra-competent *E. coli* cells for large scale production of library plasmid.

were incubated at 50˚C for 30 minutes on an Isotemp™ Incubator. Reaction mixtures were used to transform One Shot™ TOP10 Electrocomp™ *E. coli* (Thermo Fisher Scientific) with very high transformation efficiency (> 3-5 x 10^9 cfu/µg). Reaction mix of 2 µL volume was used to transform 50 µL of bacterial cells at 2500 V on a MicroPulser™ Electroporation system (Bio-Rad). After electroporation, 950 µL of SOC medium was added to each transformation. Ten transformations were performed for each 20 µL reaction mix and 30 transformations were performed for all three reaction mixtures. Transformation mixtures (~ 1 mL) for one reaction mix were pooled together after adding SOC medium and transformations for different reaction mixtures were kept separate for an hour of incubation at 37ºC with shaking at 200 rpm on Innova® 44/44R Stackable Incubator Shakers (New Brunswick Scientific). The transformation efficiency was assessed by taking 10 µL from each pooled transformation mixture and serially diluting and plating on 2 X YT/carb (2 X yeast extract and tryptone media with 100 µg/ml carbeinicillin) agar plates (Table 3. 4). Colony numbers were counted to determine bacterial cell numbers, which were calculated 2.2 x 10^6, 2.5 x 10^6, and 0.64 x 10^6 for three reaction

Table 3.4 Colony numbers from different transformation mixtures at different dilutions

Transformation mix	**Colony number: for x 10^4 dilution**	**Colony number: x 10^5 dilution**	**Colony number: for x 10^6 dilution**	**Colony number: for x 10^7 dilution**	**Colony number: for x 10^8 dilution**
Mix 1	> 100	**22**	1	0	0
Mix 2	> 100	**25**	1	0	0
Mix 3	**64**	4	1	1	0

mixtures. In total, this gave 5.34 x 10^6 cells, covering the circRNA shRNA library > 350 times. Thirty pooled transformation mixtures were collected together (30 mL) and 15 mL was transferred to two sets of 1.5 L 2YT/carb medium to incubate at 37ºC overnight with shaking at 200 rpm on Innova® 44/44R Stackable Incubator Shakers. The circRNA shRNA library plasmid was isolated from the culture using Hispeed DNA Plasmid Maxi-prep Kit (QIAGEN), according to the manufacturer's protocol. DNA concentration and purity were quantified using NanoDrop™. Library plasmid was aliquoted and stored at -20ºC.

3.2.5 Validation of the CircRNA shRNA Library

The circRNA shRNA library was first validated using *Xho*I restriction digestion to determine the corresponding DNA band sizes. CircRNA shRNA library plasmid (1.5 µg) was mixed with 2 µL of 10 X FastDigest Buffer and 2 µL of FastDigest *Xho*I, and DNAase-free ddH_2O was added to 20 µL. pLKO.1 was digested using the same reaction conditions as a control. Both reactions were incubated at 37˚C for 30 minutes on an Isotemp™ Incubator. The digested product was mixed with Gel Loading Dye (4 µL) and resolved on a 2% low-melting UltraPure™ Agarose gel at 80 V for 1 hour in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1X TAE. Expected DNA bands were detected on Gel Doc™ XR+ Imager using 1 kb DNA Ladder.

NGS method was used to analyze individual circRNA shRNA library sequences and their relative abundance. Conditions were optimized to conduct a PCR amplification using 213 ng/µL circRNA shRNA library plasmid (27.8 x 10^9 copies of oligonucleotides in 1 µL of library plasmid). For the PCR, 10 µL of 5 X HF Buffer, 1 µL of 10 mM dNTPs, 2.5 µL of 10 µM Forward Primer, 2.5 µL of 10 µM Reverse Primer and 0.5 µL of Phusion High-Fidelity

DNA Polymerase were mixed with 0.5 µL of (13.9×10^9 copies of oligonucleotides, > 9.1×10^5 times coverage of the library). DNAase-free ddH_2O was added to a final volume to 50 µL. PCR was performed with a Thermal Cycler using the following temperature profile: 1 min at 98˚C and 28 cycles of amplification (1 min at 98˚C, 15 sec at 60˚C, 1 min at 72˚C) and 1 min at 72˚C. An additional reaction mix of 50 µL was prepared using 10 µL of 5 X HF Buffer, 2 µL of 10 mM dNTPs, 10 µL of 10 µM Forward Primer (oligonucleotide 1 in Table 3.5) and 10 µL of 10 uM Reverse Primer (oligonucleotide 2 in Table 3.5), 1 µL of Phusion High-Fidelity DNA Polymerase and 17 µL of DNAase-free ddH_2O. A reaction was prepared by mixing 25 µL of first PCR reaction and 25 µL of additional reaction mix (25 µL of first PCR reaction was kept for checking the size). An extra PCR cycle was run as, 2 min at 98ºC, 20 sec at 60ºC and 2 min at 72ºC. A fraction (~ 10 µL) of both first and second PCR products was mixed with 2 µL of 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel run at 80 V for 50 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 205 bp expected band was detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder. The 40 µL product from the second PCR was mixed with 8

Table 3.5 Oligonucleotides for shRNA library preparation from genomic DNA (IDT)

ID	Oligonucleotide	Sequence (5′-3′)
1.	Ion Torrent F.P. (A primer)	*CCATCTCATCCCTGCGTGTCTCCGAC*[1]TCAG**CTAAGGTAAC** CTTTATATATCTTGTGGAAAGGACGAAACA[1]
2.	Ion Torrent R.P. (tRP1 primer)	*CCTCTCTATGGGCAGTCGGTGAT*[2]ATGAATACTGCCATTTG TCTCGACGTC[2]
3.	F.P. for genomic DNA PCR	GAGGGCCTATTTCCCATGATTC
4.	R.P. for genomic DNA PCR	GGTGGTGTGGTGTAAGG
5.	Oligo 1 for *Sal*I adapter	TCGACCTCGAGACAAATGGCAGTATTC
6.	Oligo 2 for *Sal*I adapter	GAATACTGCCATTTGTCTCGAGG

F.P.: Forward primer, R.P.: Reverse primer, *A sequence*[1], Key, **Barcode (one example, eight different barcodes were used for eight different genomic DNA samples)**, Framework[1], *P1 sequence*[2], Complementary to adapter[2]

μL of 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel at 80 V for 50 minutes in a Mini-Sub® Cell GT Systems. Agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 205 bp expected band was excised on a UV Transilluminator. The expected product was purified by QIAquick Gel Extraction Kit according to the manufacturer's protocol and eluted in 50 μL DNAase-free ddH_2O. DNA concentration and purity were quantified using NanoDrop™. Library quality assessment, Ion Torrent sequencing, data processing and analysis are discussed in sections 3.5.9 and 3.5.10.

3.3 Cell Lines and Culture Conditions

The HEK-293T cell line was used for generating lentivirus of circRNA shRNA library. HCT 116 (colorectal carcinoma) cell line was used for pooled screening with the lentiviral library. HCT 116, DLD-1 (colorectal adenocarcinoma), MDA-MB-231 (breast adenocarcinoma), MCF7 (breast adenocarcinoma), Du 145 (prostate carcinoma), and PC-3 (prostate adenocarcinoma) cell lines were used for validation experiments. Cell lines were purchased from the American Type Culture Collection (ATCC). Cells were passaged for less than three months following resuscitations and therefore no additional authentication was performed. Monolayer cultures of HEK-293T and MCF7 cell lines were maintained in the DMEM medium (HyClone, GE Life Sciences) containing 10% FBS (Gibco, Life Technologies) and 1% penicillin/streptomycin (Gibco, Life Technologies). MDA-MB-231 cell line was maintained in the DMEM medium containing 10% FBS, 1% penicillin/streptomycin, and 1 mM sodium pyruvate (HyClone, GE Life Sciences). HCT 116 and DLD-1 cell lines were maintained in McCoy's 5A medium containing 10% FBS and 1% penicillin/streptomycin. Du 145 and PC-3 cell lines were maintained in RPMI (HyClone, GE Life Sciences) containing 10% FBS, 1% penicillin/streptomycin. ATCC guidelines were followed during storing, seeding, growing, splitting, washing (e.g. 1 X PBS) and harvesting (e.g. 0.25% trypsinaization) of the cell lines. Standard incubation conditions (37ºC and 5% CO_2) were also followed for culturing the mentioned cell lines. Cell confluency was maintained between 70-80% and monitored using a light microscope. A hemocytometer was used to count the cells at different steps of the experiments.

3.4 Generation of Lentiviral Particles

CircRNA shRNA library lentivirus was generated for pooled screening. 3.5 x 10^6 HEK-293T cells were seeded in a 10 cm tissue culture dish with 10 mL of DMEM medium. After 24 hours of incubation, the culture medium was replaced with 10 mL of new DMEM medium. The following three plasmids were pooled to prepare the transfection mix: 6000 ng of circRNA shRNA library plasmid, 5400 ng of psPAX2 plasmid (virus packaging), and 600 ng of pMD2.G plasmid (virus envelop). A mixture containing 540 µL of Opti-MEM medium and 36 µL of 1 mg/mL polyethyleneimine (PEI) was added drop wise to plasmid mixture and incubated for 30 minutes at room temperature. Transfection reagent was mixed and added dropwise on to a 10 cm cell culture plate seeded with HEK-293T cells. After 18 hours of incubation, cell culture medium with transfection reagent was removed and replaced with new 10 mL of DMEM medium, containing 20% BSA. Cells were incubated for 24 hours again. Lentivirus containing media was harvested and stored at 4ºC. Another 10 mL of DMEM medium containing 20% BSA was added to the transfected cells and incubated for 24 hours. Viral harvesting was repeated and combined with the first viral harvest. Lentivirus containing medium was centrifuged at 500 X g for 5 minutes to remove any transfected cells or debris collected during harvesting. Multiple transfections were performed to generate enough lentivirus for future use, which were pooled together (20 mL of virus/dish), aliquoted, and stored at -80ºC. The lentivirus generation method is outlined in Fig 3.4.

3.5 Pooled CircRNA shRNA Screening in Cell Lines

3.5.1 Determination of Multiplicity of Infection

Cell doubling time, puromycin (Thermo Fisher Scientific) and polybrene (Sigma-Aldrich) sensitivities of HCT 116 cells were measured before pooled screening with the circRNA shRNA library. Viral titer was also assessed to determine 'Multiplicity of Infection' (MOI) for transducing HCT 116 cell line. HCT 116 cells were transduced with a very low MOI (0.3 to 0.4) to ensure that the cells were transduced by only one viral particle. For determining MOI, 3 x 10^6 of HCT 116 cells were transduced in duplicates with 0, 1, 2, 3, 4 and 5 mL of the lentiviral library respectively. Each transduction was done in a T175 tissue culture flask with required McCoy's 5A medium to a final volume of 30 mL containing 8 µg/mL polybrene. A control was set in duplicate by seeding 3 x 10^6 of HCT 116 cells in T175 tissue culture flask

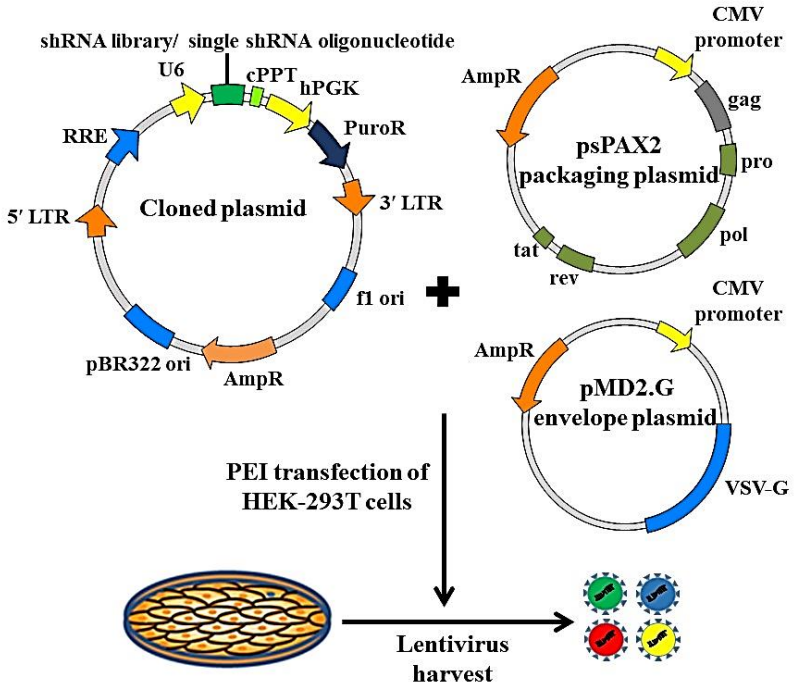

Figure 3.4 Generation of lentivirus from library plasmid. Illustration showing generation of lentiviral particles from cloned library plasmid. HEK-293T cells were co-transfected with plasmid library, viral packaging plasmid and envelope plasmid through PEI transfection method. This lentiviral library was used for pooled screening in HCT 116 cell line.

with the same culture medium to a final volume of 30 mL. After 24 hours of incubation, the culture medium was replaced with 30 mL of culture medium containing 2 µg/mL puromycin (final concentration) for each virus treatment. For the control, 30 mL of McCoy's 5A medium without puromycin was added to each transduction. After 48 hours of incubation, cells from all replicates of treatment and control groups were collected and counted. To determine the percentage of cell survival, the number of living cells from each volume of virus treatment was compared to the average number of living cells from the duplicates of control. Thus, the percentage of cell survival indicated the percentage of successfully transduced cells that acquired puromycin resistance. As 30-40% of cell survival was considered MOI of 0.3-0.4, appropriate viral volume resulting this MOI was determined. From MOI graph, 2 mL of virus was calculated for 30-40% of cell survival representing the expected MOI (Fig. 3.5).

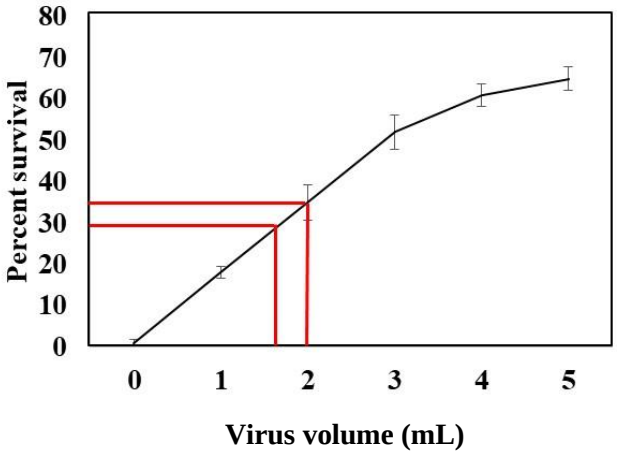

Figure 3.5 Determination of Multiplicity of Infection. Line graph showing MOI calculation for different virus titers in HCT 116 cell line. Transduction of HCT 116 cells with different viral titers (0-5 mL) was performed for 24 hours. The transduced cells underwent puromycin selection for 48 hours. After selection, 2 mL of virus showed ~ 33% cell survival. The respective MOI was calculated as 0.33 from the graph. This virus titer was used to transduce HCT 116 cell line for pooled screening.

3.5.2 Transduction of HCT 116 Cell Line with Lentivirus of CircRNA shRNA Library

To perform primary screening, 120 x 10^6 HCT 116 cells were seeded in T175 tissue culture flasks (40 flasks, ~ 3 x 10^6 cells/ flask) with 2 mL of lentivirus (MOI of 0.3-0.4) in each flask. McCoy's 5A medium was added to a final volume of 30 mL containing 8 µg/mL polybrene. After 24 hours of transduction, cells were washed with warm PBS and culture medium was replaced by 30 mL of the McCoy's 5A medium, containing 2 µg/mL puromycin. After 48 hours of puromycin treatment, culture medium was removed and cells were washed with warm PBS. Cells were then harvested from all plates and pooled together. Two replicates of 10 x 10^6 cells were collected as time point 0 (T0) samples and centrifuged at 300 x g for 6 minutes. Each replicate represents the library > 500 times, assuming that most of the cells were transduced with only one viral particle at low MOI (0.3-0.4). The cell pellet was store at -80ºC until genomic DNA extraction. Three replicates of 9 x 10^6 cells were passaged where each replicate was a set of three flasks containing 3 x 10^6 cells each. The lentiviral transduction method is outlined in Fig 3.6.

3.5.3 Outgrowth Assay

Every 2-3 days, 10 x 10^6 transduced cells were pelleted from each replicate and stored at -80ºC and 9 x 10^6 cells were passaged again. Thus, cells were passaged and pelleted for 14 generations. Cells were maintained in DMEM containing 2 µg/mL puromycin throughout the outgrowth assay. The outgrowth assay is outlined in Fig. 3.6.

3.5.4 PCR Amplification of shRNA Library from Genomic DNA

Genomic DNA was isolated from cells using the QIAamp DNA Blood Maxi Kit (QIAGEN), according to manufacturer's protocol. DNA concentration and purity were

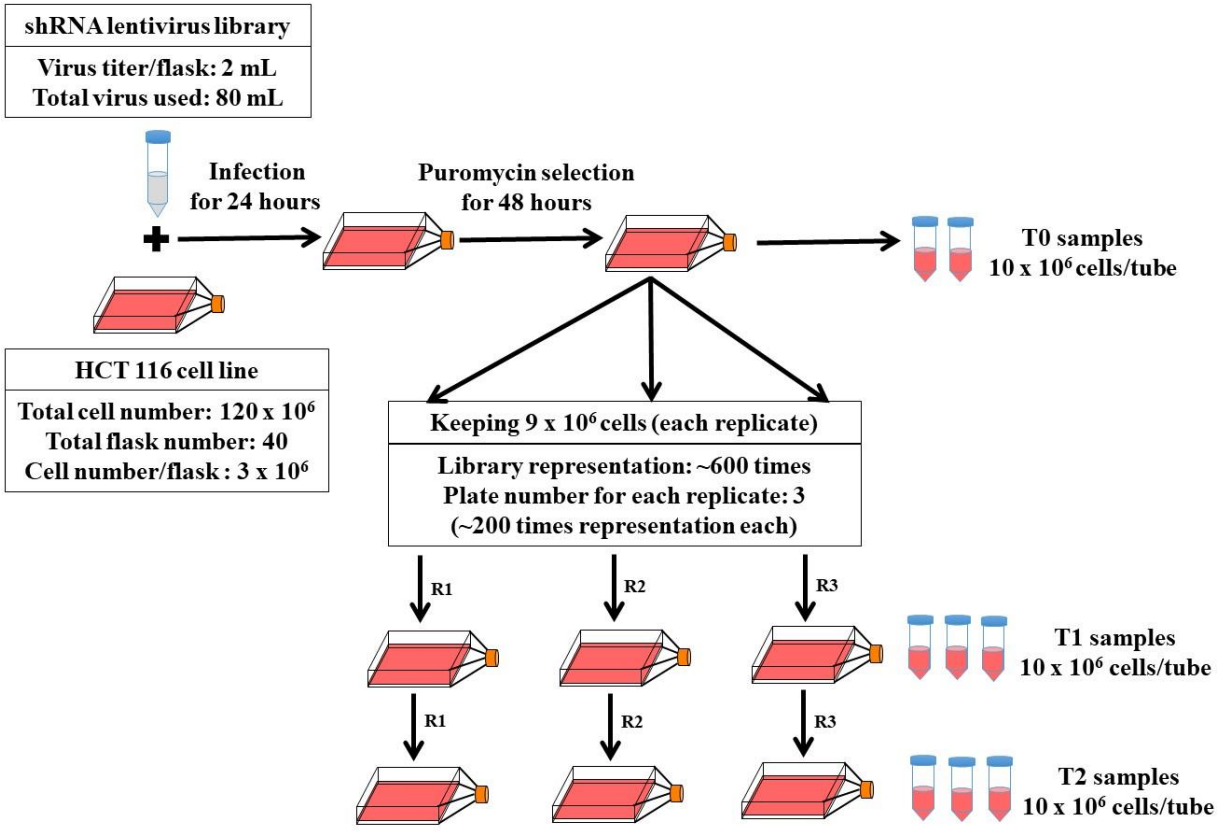

Figure 3.6 Transduction of HCT 116 cell line with lentiviral library. Illustration showing transduction procedure with lentiviral library. Cells were transduced with the lentiviral library for 24 hours according to the calculated MOI. The transduced cells underwent puromycin selection for 48 hours. Selected cells were passaged and collected up to several generations. Initial cell pellets after puromycin selection was considered T0 samples. Cell pellets from passaging of T0 cells were considered T1 samples after seven generations and T2 after fourteen generations.

quantified using NanoDrop™ 2000/2000c and the DNA stored at -20ºC. Genomic DNA was extracted from the cell pellets of the initial time point (T0), a middle time point (T1) and the last time point (T2) of the screening procedure. Sections 3.5.4 to 3.5.10 are discussed for a replicate of T0 genomic DNA sample. The same methods were applied to all other genomic DNA samples as described by Islam *et. al.,* 2017. Primers (Oligonucleotides 1 and 2 (Table 3.5) were used for PCR amplification of the circRNA shRNA library from genomic DNA. For half-shRNA sequencing, the library was PCR-amplified using oligonucleotides 3 and 4 (Table 3.5) to generate a larger amplicon. As a result, in the next step, digestion with *Xho*I resulted in a 316 bp digestion product, which was easily separated from the smaller 43 bp digestion product and removed during PCR clean up. PCR was performed using the following reaction: 160 µL of 10 X Pfx Amplification Buffer (Invitrogen), 160 µL of 10 X PCR_X Enhancer Solution (Invitrogen), 24 µL of 10 mM dNTPs, 30 µL of 25 µM Forward Primer, 30 µL of 25 µM Reverse Primer, 24 µL of 50 mM $MgSO_4$ (Invitrogen) and 12 µL of Platinum™ Pfx DNA Polymerase (Invitrogen) were added together with 20 µg of genomic DNA template. DNAase-free ddH_2O was added to give a final volume of 800 µL. The 800 µL PCR was divided into 16 reactions with each reaction as a 50 µL aliquot and PCR amplification was performed with a Thermal Cycler. The temperature profile for the PCR was set as 3 min at 98˚C, 30 cycles of amplification (10 sec at 98˚C, 15 sec at 55˚C, 15 sec at 72˚C) and 5 min at 72˚C. We used 20 µg template to ensure enough representation of each sequence of the circRNA shRNA library during the amplification. PCR reactions were pooled together and purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific), according to manufacturer's protocol and eluted in 100 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using NanoDrop™. About 10 µL of the PCR product was mixed with 2 µL of 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel at 90 V for 45 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 205 bp (shRNA sequencing) and 359 bp (half-shRNA sequencing) expected bands were detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder. For shRNA sequencing, 2 µg of the 205 bp product was mixed with 6 X DNA Gel Loading Dye and resolved on a 2% low-melting UltraPure™ Agarose the gel at 90 V for 45 minutes in a Mini-Sub® Cell GT Systems. The gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 205 bp expected band was excised on a UV Transilluminator. The expected product was purified by QIAquick Gel

Extraction Kit according to the manufacturer's protocol and eluted in 25 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using NanoDrop™. The 359 bp product was used in the next step for making half-shRNA.

3.5.5 Removing Hairpin from shRNA by Restriction Digestion

Half of the hairpin present in shRNAs (half-shRNA) was removed by digesting the purified PCR product (circRNA shRNA library from genomic DNA) with *Xho*I. To perform the *Xho*I digestion reaction, 12 µg of PCR product was mixed with 120 µL of 10 X FastDigest Buffer, 60 µL of FastDigest *Xho*I (Thermo Fisher Scientific), and DNAase-free ddH_2O to bring the volume to 1.8 mL reaction. The entire 1.8 mL reaction was divided into 10 aliquots of 180 µL. Reactions were incubated at 37˚C for 15 minutes on an Isotemp™ Incubator. The *Xho*I restriction enzyme reaction was performed immediately after PCR amplification to avoid cruciform formation. Digestion reactions were pooled together and purified with GeneJET PCR Purification Kit according to the manufacturer's protocol. PCR Purification Kit removed the smaller *Xho*I-digested product at this stage and larger *Xho*I-digested product was eluted in 20 µL DNAase-free ddH_2O. Purified and digested PCR product was mixed with 4 µL of 6 X DNA Gel Loading Dye and resolved on a 2% low-melting UltraPure™ Agarose gel at 90 V for 45 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 316 bp expected band was detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder (some undigested product was also found at 359 bp). The 316 bp band was cut on a UV Transilluminator. The *Xho*I-digested product was purified by QIAquick Gel Extraction Kit according to the manufacturer's protocol and eluted in 20 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using a NanoDrop™.

3.5.6 Preparation and Ligation of an Adapter to Half-shRNA

A *Sal*I adapter was ligated to the half-shRNA. The *Sal*I adapter was prepared using oligonucleotides 5 and 6 (Table 3.5). This adapter sequence has a common complementary sequence at 3′-end for hybridizing to the Ion Torrent R.P. primer. To prepare the adapter, 10 µL of 200 µM Oligo 1 *Sal*I adapter and 10 µL 200 µM Oligo 2 *Sal*I adapter were added to 30 µL DNAase-free ddH_2O to make the 50 µL annealing reaction. The reaction was heated at 98˚C, 80˚C, 70˚C, 60˚C, 55˚C, 50˚C, 40˚C, and 25˚C for 1 minute at each temperature with a

Thermal Cycler and 950 µL of DNAase-free ddH_2O was added. DNA concentration and purity were quantified using NanoDrop™. *Sal*I adapter was ligated to *Xho*I-digested half-shRNA product. The *Sal*I adapter contains a cohesive compatible end with the *Xho*I-digested product and common complementary sequence for hybridizing to the Ion Torrent R.P. primer. A 1: 3 molar ratio of *Xho*I-digested half-shRNA product and *Sal*I adapter was used. For this purpose, 600 ng of *Xho*I-digested PCR product was mixed with 5 µL of the *Sal*I adapter (42.7 ng/µL). To perform the ligation reaction, 5 µL of T4 DNA Ligase (Invitrogen), 200 µL of 5 X Ligase Reaction Buffer (Invitrogen) and DNAase-free ddH_2O was added to give a final volume of 1 mL. The reaction was split into 10 aliquots of 100 µL and incubated at 25˚C for 1 hour using an Isotemp^{TM} Incubator. Ligation reactions were pooled together and *Sal*I adapter-ligated product was purified with the GeneJET PCR Purification Kit according to the manufacturer's protocol, which removes *Sal*I self-ligated product as well. Ligated product was eluted in 25 µL DNAase-free ddH_2O and DNA concentration and purity were quantified using a NanoDrop™.

3.5.7 Barcode Labeling of Adapter-ligated Half-shRNA Product

Primers 1 and 2 (Table 3.5) with a barcoding sequence were attached to the ligated product. Eight different barcodes (designed according to Ion Torrent sequencing guidelines) were used for eight different genomic DNA samples. For the PCR, 30 µL of 10 µM Ion Torrent Barcode Forward Primer, 30 µL of 10 µM Ion Torrent Barcode Reverse Primer and 250 µL of 2 X Phusion Master Mix with HF Buffer were assembled with 50 ng ligated product and DNAase-free ddH_2O was added to give a final volume of 500 µL. The reaction was divided into 10 aliquots of 50 µL and PCR amplification was performed with a Thermal Cycler. The temperature profile for the PCR was 30 sec at 98˚C, 28 cycles of amplification (10 sec at 98˚C, 5 sec at 56˚C, 5 sec at 72˚C) and 15 sec at 72˚C. PCR reactions were pooled together and barcoded product was purified with the GeneJET PCR Purification Kit. Barcoded product was eluted in 25 µL DNAase-free ddH_2O and DNA concentration and purity were quantified using a NanoDrop™. It is important to optimize the PCR conditions at this step to reduce heteroduplex formation. PCR cycles were reduced from 28 cycles to 15 cycles to eliminate heteroduplex formation after optimizing the PCR cycles. As 28 cycles formed heteroduplex, PCR reactions were performed with 15 cycles for improved half-shRNA sequencing.

3.5.8 Removal of Self-ligated Product

To eliminate the *Xho*I self-ligated product, *Xho*I digestion was performed. Briefly, 1 µg of barcoded product was mixed with 10 µL of 10 X FastDigest Buffer and 5 µL of FastDigest *Xho*I and the final volume was increased to 150 µL using DNAase-free ddH_2O. Digestion reaction was incubated at 37˚C for 15 minutes with an Isotemp™ Incubator. After the removal of the self-ligated products, the barcoded PCR product was purified with the GeneJET PCR Purification Kit according to the manufacturer's protocol and DNA was eluted in 50 µL of DNAase-free ddH_2O. For the lower cycle barcoded product (15 cycles), magnetic bead-based purification was done instead of gel extraction to reduce sample loss. In this case, the eluted product was purified with Agencourt® AMPure® XP Reagent (Beckman Coulter) according to manufacturer's protocol. This was a two-step purification method where the first step removed all the high molecular weight DNA contamination (e.g. genomic DNA) and the second step removed all the smaller DNA fragments (e.g. primer, primer dimer, and restriction digested fragments). In both processes, DNA concentration and purity were quantified using NanoDrop™.

3.5.9 Quality Assessment of Library

Throughout all procedures described above, the quality of the library sample or any PCR product was assessed with the Agilent 2100 Bioanalyzer, using the Agilent High Sensitivity DNA chip, according to the manufacturer's protocol. For this purpose, 1 µL of 500 pg/µL sample was applied to the chip. A sharp peak was expected for a pure library sample. Presence of multiple peaks and/or broad peak larger than the expected library suggested formation of a heteroduplex. Electropherograms for this analysis were generated with 2100 Expert Software.

3.5.10 Ion Torrent Sequencing, Data Processing and Analysis

Though this method is explained for one sample, multiplexing of different samples with specific barcodes (barcode labeling) was performed to sequence multiple samples at the same time. Amplicon concentration was determined using a NanoDrop™ and 25 µL of DNA (26 pM) was prepared for emulsion PCR. Emulsion PCR was performed using the Ion PGM™ Hi-Q™ OT2 Kit (Life Technologies), according to manufacturer's protocol. First, a unique DNA amplicon was amplified and bound to a single Ion Sphere Particle (ISP) by emulsion PCR.

Amplification primers that bind to A and P1 adapters were used for clonal amplification so that each ISP was covered with many copies of the same DNA fragment. Second, because the A primer was biotinylated, template positive ISPs could be isolated using Ion Torrent enrichment beads and non-templated ISPs were removed. Third, dsDNA anchored to the ISPs was denatured. This allowed the ISPs with ssDNA to go into solution while the biotinylated strand remain bound to enrichment beads. The solution containing ssDNA enriched ISPs was used for Ion Torrent sequencing. This NGS method was performed with Ion PGM™ System (Thermo Fisher Scientific). Ion 318 chips and Ion PGM Hi-Q Sequencing Kits were used according to manufacturer's protocol in this regard. Base calling, chip analysis, and barcode separation were performed using the Ion Torrent Server Software 5.0 package. Chip analysis included percentage ISP loaded, percentage of enriched ISPs, percentage of polyclonal reads (ISPs with multi-type DNA templates), and percentage of low quality reads. Total raw sequence reads were retrieved as FASTQ format downloaded from Ion Torrent server, counted, compared to the designed library sequences and plotted accordingly. All the scripts for the analysis of sequencing results were written on Python-2.7.9. Each read was scanned for an established known set of sequence (represented as framework in Table 3.5) followed by the number of nucleotides equal to the length of the library sequences (in our case, 21) and then the cleaved-*Xho*I site ('CTC'). Once all reads had been scanned, count of each unique library sequence was saved to a file. The data processing and analysis methods are outlined in Fig. 3.7.

3.5.11 Sequencing Library Preparation for GeCKO Library

For quality assessment of the GeCKO library from plasmids, two-step PCR amplification was performed. For the first PCR (PCR 1) using naïve plasmid DNA library, 20 µL of 10 X Pfx Amplification Buffer, 20 µL of 10 X PCR_X Enhancer Solution, 3 µL of 10 mM dNTPs, 4.5 µL of 20 uM Primer Mix (using oligonucleotide 1 and 2 from Table 3.6), 3 µL of 50 mM $MgSO_4$ and 1.5 µL of Platinum™ Pfx DNA polymerase were added together with 40 ng of GeCKO library. DNAase-free ddH_2O was added to give a final volume of 100 µL. The reaction was divided in 50 µL aliquots and PCR amplification was performed with a Thermal Cycler. The temperature profile for the PCR was 5 min at 98°C, 30 cycles (note that this cycle number was reduced as described below) of amplification (15 sec at 98°C, 15 sec at 65°C, 40 sec at 72°C), and 5 min at 72°C. The same reaction was performed with multiple dilutions of

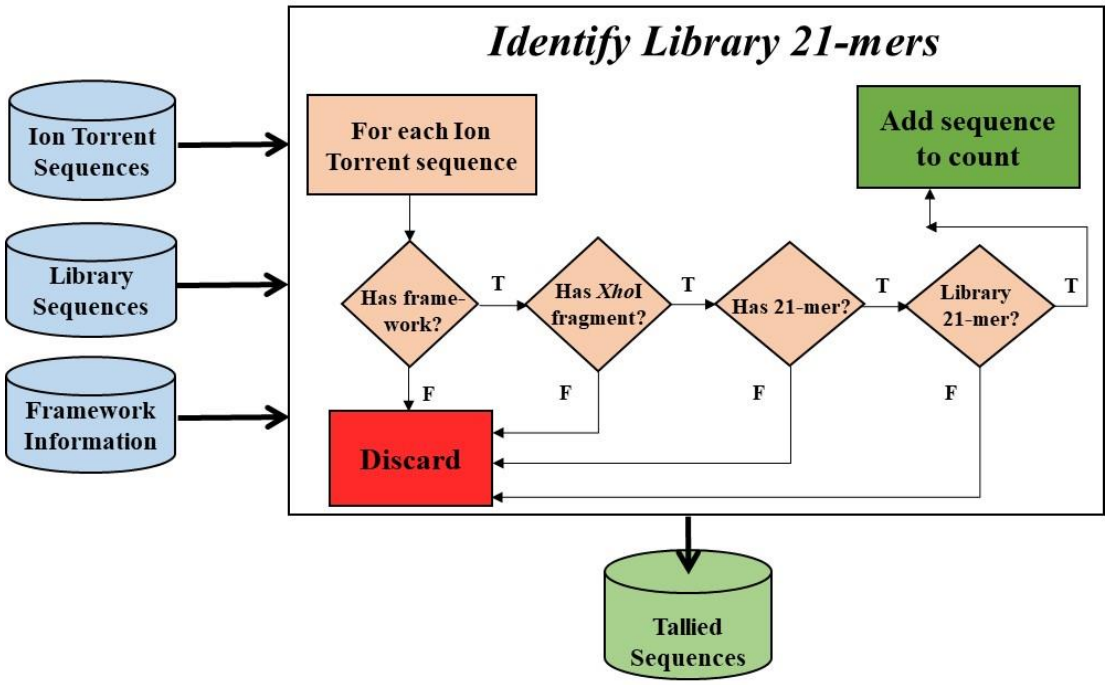

Figure 3.7 Data analysis of Ion Torrent sequencing. Flowchart showing analysis of sequencing data from NGS platform. Analysis was done by comparing raw sequence reads from shRNA and half-shRNA library sequencing to the designed library sequences. In both cases, 21-mer sequences were considered for this comparison. Eventually, the respective counts of the sequence reads were calculated.

the same library using different barcodes. The reactions were pooled together. About 10 µL of the PCR product was mixed with 2 µL of 6 X DNA Gel Loading Dye and resolved on a 2% low-melting UltraPure™ Agarose gel at 100 V for 1 hour in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The expected 312 bp band was visualized using 50 bp DNA Ladder on a UV Transilluminator. For the second PCR (PCR 2) 5 µL of PCR 1 product was used as a template. Briefly, 20 µL of 10 X Pfx Amplification Buffer, 20 µL of 10 X PCRx Enhancer Solution, 3 µL of 10 mM dNTPs, 4.5 µL 20 µM Primer mix (using oligonucleotide 3 and 4 from Table 3.6), 2 µL of 50 mM $MgSO_4$ and 1.5 µL of Platinum™ Pfx DNA Polymerase were added together with 10 µL of amplicon from PCR 1 (for a 2 X reaction) and DNAase-free ddH_2O was added to give a final volume of 100 µL. The reaction was divided in 50 µL aliquots and PCR amplification was performed with a Thermal Cycler. Optimization of PCR annealing temperature and cycle number for Illumina primers was carried out. The temperature profile for the PCR was 5 min at 94˚C, 25 cycles (note that this cycle number was reduced as described below) of amplification (15 sec at 94˚C,

30 sec at 63˚C, 23 sec at 72˚C), and 5 min at 72˚C. The reactions were pooled together and mixed with 20 µL of 6 X DNA Gel Loading Dye and resolved on a 2% low-melting UltraPure™ Agarose gel at 100 V for 1 hour in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The expected 370 bp band was excised, using 100 bp DNA Ladder as a guide, on a UV Transilluminator. QIAquick Gel Extraction Kit was used to purify the library according to the manufacturer's protocol. DNA was eluted in 30 µL DNAase-free ddH_2O and DNA concentration and purity were quantified using a NanoDrop™. The quality of samples was assessed on Bioanalyzer as previously described for shRNA library in section 3.5.9. Since Bioanalyzer data showed formation of heteroduplex structures, we reduced the PCR cycles. Specifically, we found that reducing first PCR cycle number to 18 cycles and the second PCR cycle number to 20 cycles eliminated heteroduplex formation. We sequenced the GeCKO library using Illumina NextSeq 500 High Output (75 Cycles, 400M Reads) according to manufacturer's protocol. Python-2.7.9 was used to generate scripts for analyzing the data. For assessment of the GeCKO library from genomic DNA, the PCR was performed as described above. However, the genomic DNA template concentration was set at 30 µg to achieve higher representation of the integrated library.

Table 3.6 Oligonucleotides for GeCKO library preparation (IDT)

ID	Oligonucleotide	Sequences (5′-3′)
1.	F.P. for genomic DNA PCR	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
2.	R.P. for genomic DNA PCR	CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC
3.	Illumina F.P.	*AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT*T**AAGTAGAG**TCTTGTGGAAAGGACGAAACACCG
4.	Illumina R.P.	*CAAGCAGAAGACGGCATACGAGAT***AAGTAGAG***GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*TTCTACTATTCTTTCCCCTGCACTGT

F.P.: Forward primer, R.P.: Reverse primer, *Illumina adapter*, Stagger, **Barcode (one example)**, Priming site

3.6 Analysis of Screen Results

All the scripts for the analysis of the screen results were written on Python-2.7.9. Raw sequence files (FASTQ) were extracted and processed to compare the sequence reads to the designed library sequences using the same method described in Fig. 3.7. In order to increase the counts of the library sequences for statistical purposes, the counts of all the sequence were summed up from four different NGS runs of a sample. This led to 1,044,443 sequences for T0 time point, 888,587 sequences for T1 time point and 1,046,616 sequences for T2 time point. The count of each library sequence was smoothed by five (the raw count of each library sequence was summed by five) and then normalized to per million of total library sequence count of a time point. The count of each library sequence was then compared between different time points (T0, T1 and T2). If a given knockdown of a circRNA by any shRNA causes lethality, then those cells with that shRNA integration gets depleted from the rest of the population (dropouts). We used a scoring system that filters the shRNA sequences that show more than 1.25 fold increase of from initial time point (T0) to middle time point (T1). The difference of cumulative change (DCC) for remaining shRNA sequences was calculated based on the following formula:

$$DCC = [\{\log_e(T1\ value) - \log_e(T0\ value)\} + \{\log_e(T2\ value) - \log_e(T1value)\}]$$

To increase the confidence in the analysis, a bootstrapping method was used to assign *p*-values based on calculated DCC scores for at least two shRNAs targeting the same circRNA. The analysis method is outlined in Fig. 3.8. The bootstrapping is a heuristic strategy which considers the highest dropout as the reference and assigns values to all other dropouts accordingly. Expression patterns of the parental genes of the top essential circRNA hits (E-circRNAs) were analyzed using 'The Cancer Genome Atlas (TCGA)' dataset (https://cancergenome.nih.gov/) and cancer cell lines from 'Cancer Cell Line Encyclopedia (CCLE)' dataset (http://www.broadinstitute.org/ccle). Some enrichment studies on the genes of top genic essential circRNA hits were performed using FunRich version 3 software packages using the default parameters. Top five essential circRNAs were selected for validation and respective sequences from our library were used to clone individual shRNA oligonucleotides.

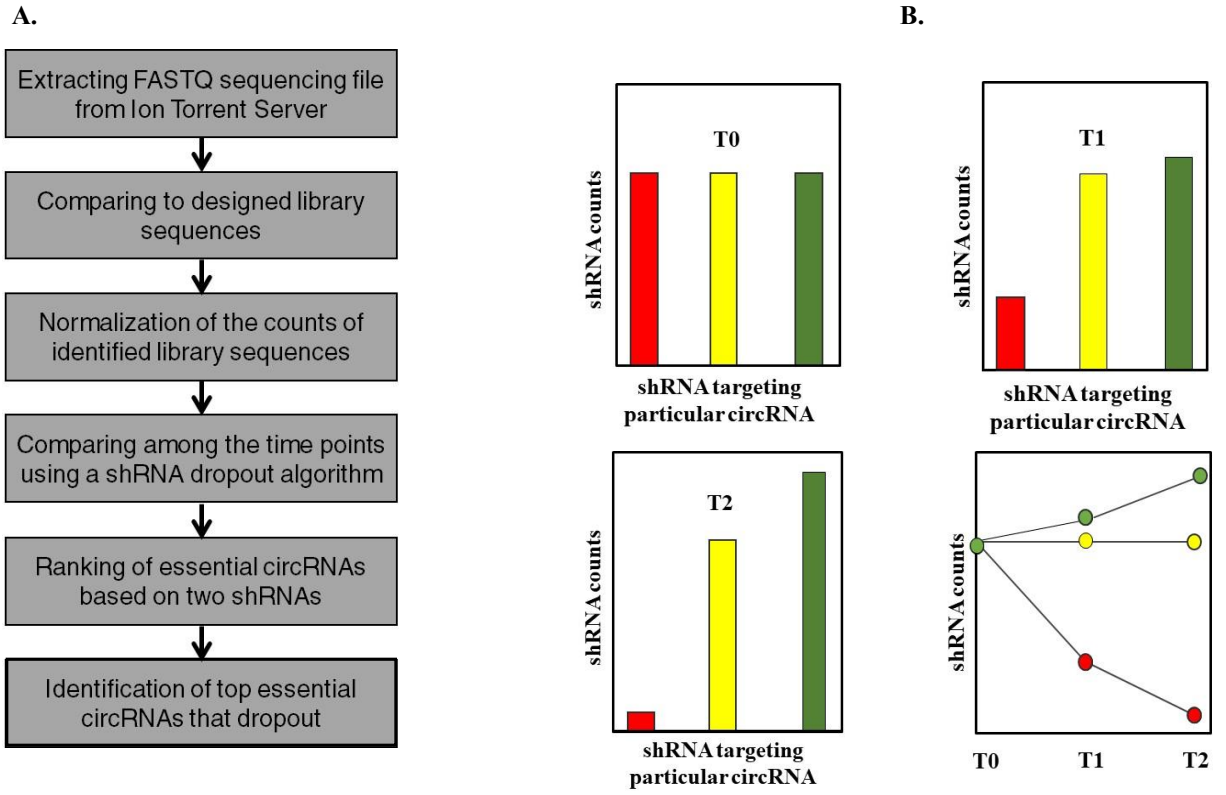

Figure 3.8 Pooled screening data analysis. (A) Flowchart showing a pipeline of data analysis methodology for pooled screening. **(B)** Diagrams showing the rationale to identify sequence dropouts in the screening procedure. The raw sequence reads were compared to the designed library sequences. The counts of the library sequences went through some normalization process. Essential circRNAs were identified by comparing the normalized counts of library sequences among T0, T1 and T2 time points. The essential circRNAs were listed based on the described scoring algorithm. Top essential circRNAs were selected for validation experiments.

3.7 Validation of Top Essential Circular RNAs from Screening

3.7.1 PCR Amplification, Cloning and Validation of shRNA Oligonucleotides

For validating the top five circRNA hits found from the screening, individual shRNA oligonucleotides targeting these circRNAs were amplified and cloned into pLKO.1-TRC plasmid (Table 3.7). Here the procedures are shown for one shRNA oligonucleotide as outlined in Fig. 3.9. To perform the PCR reaction, 20 µL of 5 X HF Buffer, 2 µL of 10 mM dNTPs, 2.5 µL of 10 uM Forward Prime (oligonucleotide 1, Table 3.1), 2.5 µL of 10 uM Reverse Primer (oligonucleotide 2, Table 3.1) and 1 µL of Phusion High-Fidelity DNA Polymerase were mixed

Table 3.7 shRNA oligonucleotides for top essential circRNAs (IDT)

Hit ID	**shRNA oligonucleotide sequence (5′-3′)**	
CircRNA Hit 1	shRNA 1	CCTCACCACTCTCCACCCAAACTTCTCAATGACTTGTCCTCGAG GACAAGTCATTGAGAAGTTTGCCTTACACCACACCACC
	shRNA 2	CCTCACCACTCTCCACCCAATGACTTGTCCAGTTCGGACTCGAG TCCGAACTGGACAAGTCATTGCCTTACACCACACCACC
CircRNA Hit 2	shRNA 1	CCTCACCACTCTCCACCCAATCATCCGTGAGCTGATGCCTCGAG GCATCAGCTCACGGATGATTGCCTTACACCACACCACC
	shRNA 2	CCTCACCACTCTCCACCATCAATCATCCGTGAGCTGATCTCGAG ATCAGCTCACGGATGATTGATCCTTACACCACACCACC
CircRNA Hit 3	shRNA 1	CCTCACCACTCTCCACCCTAAACAGTGTGGCCACAGACCTCGA GGTCTGTGGCCACACTGTTTAGCCTTACACCACACCACC
	shRNA 2	CCTCACCACTCTCCACCTACTAAACAGTGTGGCCACAGCTCGAG CTGTGGCCACACTGTTTAGTACCTTACACCACACCACC
CircRNA Hit 4	shRNA 1	CCTCACCACTCTCCACCAAATTTCTCCCAGTGGTGACTCTCGAG AGTCACCACTGGGAGAAATTTCCTTACACCACACCACC
	shRNA 2	CCTCACCACTCTCCACCATTTGCCATGGACAAGATTTCCTCGAG GAAATCTTGTCCATGGCAAATCCTTACACCACACCACC
CircRNA Hit 5	shRNA 1	CCTCACCACTCTCCACCATGTAATCCTGGTTCACACGGCTCGAG CCGTGTGAACCAGGATTACATCCTTACACCACACCACC
	shRNA 2	CCTCACCACTCTCCACCTGTAATCCTGGTTCACACGGCCTCGAG GCCGTGTGAACCAGGATTACACCTTACACCACACCACC

with 2 µL of 20 µM shRNA oligonucleotides. DNAase-free ddH_2O was added to bring the final volume to 100 µL. The 100 µL PCR was divided into two reactions and PCR was performed with a Thermal Cycler. The temperature profile for the PCR was set as 30 sec at 98˚C and 28 cycles of amplification (5 sec at 98˚C, 10 sec at 60˚C, 5 sec at 72˚C) and 30 sec at 72˚C. Reactions were pooled together afterwards. About 10 µL of the PCR product was mixed with 2 µL of 6 X DNA Gel Loading Dye and resolved on a using a 2% UltraPure™ Agarose gel at 80 V for 50 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 142 bp expected band was detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder. The remaining PCR product was resolved (with 6 X DNA Gel Loading Dye) at 80 V for 50 minutes in a Mini-Sub® Cell GT Systems using 2% low-melting UltraPure™ Agarose gel and stained with SYBR® Safe

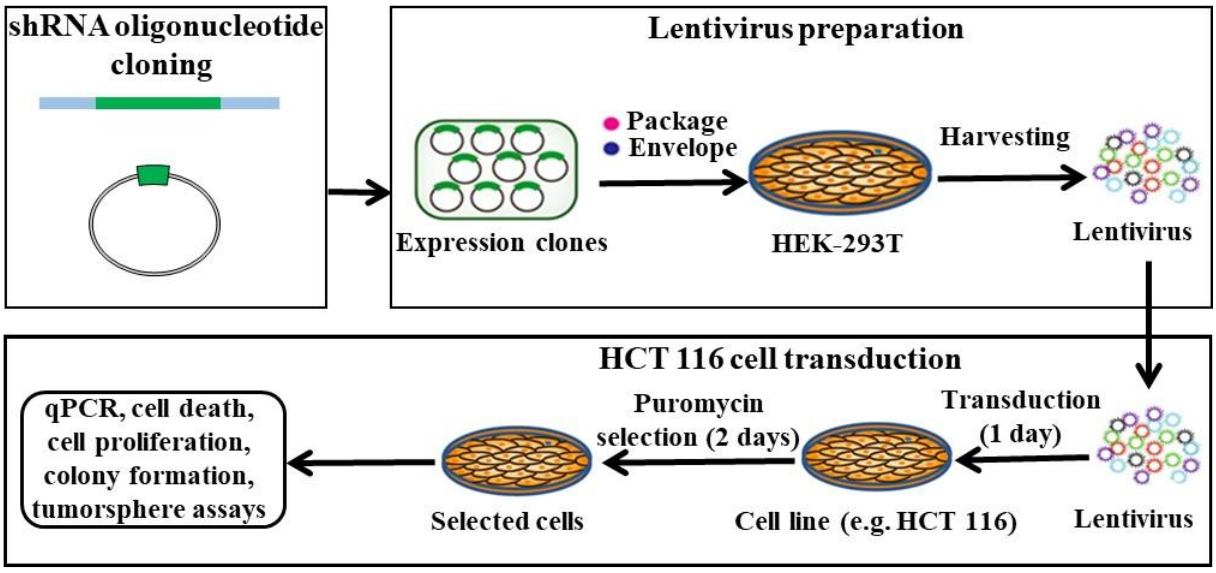

Figure 3.9 Validation of Essential circRNAs. Schematic showing validation experiments for top essential circRNA. Individual shRNA oligonucleotides were cloned into pLKO.1 plasmid. The plasmid library was packaged into lentivirus. The lentiviral library was used to transduce target cells for 24 hours. The transduced cells underwent puromycin selection for 48 hours. The selected cells were used for expression assay and several cell based experiments.

DNA Gel Stain in 1 X TAE. The 142 bp expected band was excised on a UV Transilluminator. The expected product was purified by QIAquick Gel Extraction Kit according to manufacturer's protocol and eluted in 50 µL DNAase-free ddH_2O. DNA concentration and purity were quantified using NanoDrop™.

In order to clone the PCR-amplified oligonucleotide into pLKO.1-TRC plasmid, the plasmid was digested with *Age*I and *Eco*RI to release the stuffer in the same method as described in section 3.2.4. To perform the cloning, 100 ng of *Age*I- and *Eco*RI- digested, purified pLKO.1 pLKO.1 (3.9×10^9 copies) and 3 ng of PCR-amplified oligonucleotide (142 bp, 1.9×10^{10} copies) were mixed to make a plasmid: insert molar ratio of 1: 5 in 15 µL volume. A reaction was prepared by combining 5 µL of digested plasmid/insert mix and 15 µL of Gibson Assembly® Master Mix (New England BioLabs) and incubated at 50˚C for 1 hour on an Isotemp™ Incubator. This reaction mixture was used to transform ElectroMAX DH10B Cells (Thermo Fisher Scientific) *E. coli* cells with very high transformation efficiency (> 3-5 x 10^9 cfu/µg). Reaction mix of 1 µL volume was used to transform 50 µL of bacterial cells at 2500 V on a MicroPulser™ Electroporation system. After electroporation, 950 µL of SOC

medium was added to the transformation. The transformation mixture of ~ 1 mL was incubated at 37 ºC with shaking at 200 rpm on Innova® 44/44R Stackable Incubator Shaker. Serial dilutions of transformation mixture were plated on 2 X YT/carb (2 X yeast extract and tryptone media with 100 µg/mL carbinicillin) and incubated at 37˚C overnight. Around 10 healthy colonies were picked from appropriate dilution (~ 50 µL) and were used in colony PCR. To perform this PCR reaction, 4 µL of 10 X PCR Buffer (QIAGEN) 0.8 µL of 10 mM dNTPs, 1.5 µL of 20 µM Forward Primer (oligonucleotide 1, Table 3.1), 1.5 µL of 20 µM Reverse Primer (oligonucleotide 2, Table 3.1) and 0.5 µL of HotStarTaq DNA 0.5 µL Polymerase (QIAGEN) were mixed with 31.7 µL DNAase-free ddH_2O for a single colony touch. The temperature profile of the Thermal Cycler was set as 15 min at 95˚C and 30 cycles of amplification (30 sec at 94˚C, 30 sec at 55˚C, 1 min at 72˚C), 5 min at 72˚C. About 10 µL of the PCR product was mixed with 2 µL 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel at 90 V for 45 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The colonies having 142 bp expected band was detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder. Plasmids from different colonies having the band were isolated using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. The plasmids were then sequenced on 3500/3500xL Genetic Analyzer (Applied Biosystems) using 10 µL of 50 ng/ µL sample. The sequence results were analyzed on Geneious R8 software package (Geneious) and colonies having exact shRNA oligonucleotide sequence was used to isolate plasmid (in larger amount) using Hispeed DNA Plasmid Maxi Kit according to the manufacturer's protocol. DNA concentration and purity were quantified using NanoDrop™. Finally, the cloned plasmid was aliquoted and stored at -20 ºC.

3.7.2 Generation of Lentiviral Particles and Transduction of Target Cell Lines

In order to generate lentivirus from the cloned plasmids, the same procedures were followed as discussed in section 3.4. Here, transduction methods are described for lentivirus generated from one cloned plasmid targeting a circRNA. For transduction of the target cells, 0.2 x 10^6 cells were seeded in a well of 6-well cell culture dish with 500 µL of thawed lentivirus. Culture medium was added to the final volume of 3 mL, containing 8 µg/mL polybrene. After 24 hours of transduction, cells were washed with warm PBS and culture

medium was replaced by 3 mL of the culture medium, containing 2 µg/mL puromycin. After 48 hours of puromycin treatment, culture medium was removed, cells were washed with warm PBS. The required number of puromycin resistant cells were seeded for next set of experiments (section 3.7.3-section 3.7.9) as outlined in Fig. 3.9. Some lentivirus for the control shRNA plasmid (sh-GFP/sh-RFP) was also prepared and used for transduction of the target cells. Control shRNAs are designed to target GFP or RFP and thus do not target any human proteins. These controls are used compare the effect of shRNAs targeting different circRNAs.

3.7.3 Circular RNA Knockdown Assay

The methods are described for lentivirus from one cloned plasmid targeting a circRNA after the procedures explained in section 3.7.2. About 1 x 10^6 of puromycin resistant cells (HCT 116) were seeded in a well of 6-well cell culture dish in 3 mL culture medium and cultured for 72 hours. Both floating and adhered cells were harvested together as cell pellet. The cell pellet was stored at -80˚C and later used for total RNA extraction. Total RNA from transduced cells was extracted using TRIzol™ Reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. The quality of extracted RNA was assessed after resolving 1 µg of RNA sample was mixed with 6 X DNA Gel Loading Dye on an 1% low-melting UltraPure™ Agarose gel at 90 V for 1 hour in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. The 28S rRNA and 18S rRNA bands were detected on Gel Doc™ XR+ Imager using 1 kb DNA Ladder. RNA concentration and purity were quantified using NanoDrop™ and stored at -20 ºC.

To conduct a two-step reverse transcription PCR, the total RNA was reverse transcribed to make cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. Briefly, 2 µL of 10 X RT Buffer, 0.8 µL of 25 X dNTP (100 mM), 2.0 µL of 10 X RT Random Primers, 1 µL of MultiScribeTM Reverse Transcriptase and 4.2 µL of DNAase-free ddH_2O were mixed (total 10 µL). This 10 µL of 2 X RT master mix was added to 2 µg of total RNA in 10 µL to make a 20 µL of reaction. To perform this reaction, the temperature profile of the Thermo Cycler was set as, 10 min at 25˚C, 120 min at 37˚C, 5 min at 85˚C and hold at 4˚C before next steps. The resulting cDNA was stored at -20˚C for later use.

Primers were designed for specifically measuring the expression of circRNA and corresponding mRNA (for E-circRNAs). 'Divergent Primers' option in 'CircInteractome' database (https://circinteractome.nia.nih.gov/index.html) was used which directed circRNA junction sequences to 'NCBI/Primer BLAST' (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). For designing mRNA specific primers, 'Circular RNA' option of 'CircInteractome' database was used to direct 'Best Transcript' sequences to 'NCBI/Primer BLAST'. For mRNAs corresponding to E-circRNAs, primers were designed at different exons from the circRNA forming ones using 'Clustal Omega' (https://www.ebi.ac.uk/Tools/msa/clustalo/) as described in Fig. 3.10. Mostly default parameters were used for both the programs although some of the parameters were changed to get suitable primers. Different set of primers were tested using RT-PCR from the cDNA to optimize the conditions. To perform the PCR reaction (second step of two-step reverse transcription), 10 µL of 10 X PCR Buffer, 2 µL of 10 mM dNTPs, 2.5 µL of 20 uM Forward Primer, 2.5 µL of 20 uM Reverse Primer and 0.5 µL of HotStarTaq DNA Polymerase (QIAGEN) were mixed with 100 ng of total RNA. DNAase-free ddH_2O was added to the final reaction volume 100 µL. The temperature profile of the Thermo Cycler was set as, 15 min at 95˚C and 30-40 cycles (based on the expression level) of amplification (30 sec at 94˚C, 30 sec at 55˚C, 1 min at 72˚C), 5 min at 72˚C. About 10 µL of the PCR product was mixed with 2 µL of 6 X DNA Gel Loading Dye and resolved on a 2% UltraPure™ Agarose gel at 90 V for 45 minutes in a Mini-Sub® Cell GT Systems. The agarose gel was stained with SYBR® Safe DNA Gel Stain in 1 X TAE. Appropriate bands were detected on Gel Doc™ XR+ Imager using 100 bp DNA Ladder. Primers having clean expected bands were used in next steps (Table 3.8).

Knockdown measurement of circRNAs and corresponding mRNAs (for E-circRNAs) was conducted based on a qPCR method (real-time) called 'Comparative Ct Method (ΔΔCT Method)' on StepOnePlus™ Real-Time PCR System (Applied Biosystems) according to the the manufacturer's protocol. Here, the procedures are described for one cDNA sample. SensiFAST SYBR Hi-ROX Kit (BIOLINE) was used to perform the qPCR according to the manufacturer's protocol after optimizing the condition. Briefly, 10 µL of 2 X SensiFAST SYBR Hi-ROX mix, 0.8 µL of 10 µM Forward Primer (Table 3.8), 0.8 µL of 10 µM reverse Primer (Table 3.8), 8.4 µL (100 ng cDNA) of cDNA sample and required DNAase-free ddH_2O were mixed to prepare a reaction of 20 µL. Three-step cycling was performed as, 2 min at

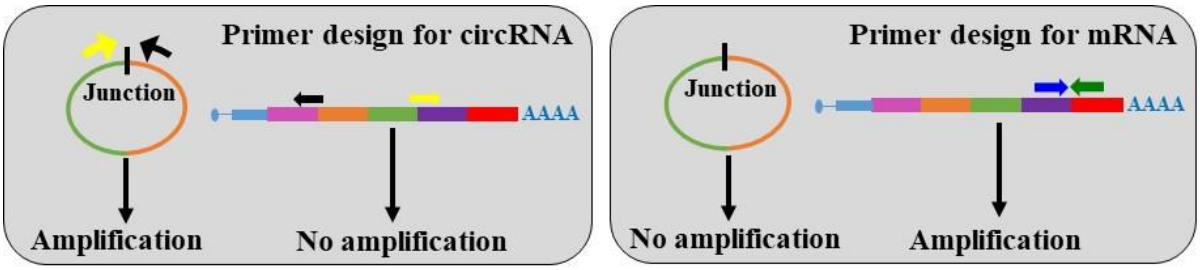

Figure 3.10 Design of qPCR primers for expression study. Schematic showing design process of primers for specific expression measurement of circRNAs and respective mRNAs (for E-circRNAs). Divergent primers were designed for circRNAs at junction regions. Convergent primers for mRNAs (linear) were designed at exons different from circRNA forming exons (for the E-circRNAs). Different colored boxes represent different exons in a mature transcript (circRNA or mRNA).

95°C (holding stage), 50 cycles of amplification (5 sec at 95°C, 10 sec at 60°C, 15sec at 72°C). A melt curve stage (step and hold) was also set for the qPCR run. StepOnePlus™ Software 2.3 package was used to set the experiment and get the Ct value, amplification plot, melt curve and other information that helped to assess the quality of qPCR run. For data analysis, Ct value of the endogenous control (GAPDH) was subtracted from Ct value of the corresponding circRNA/mRNA to get ΔCt value. To calculate –ΔΔCt value, ΔCt value of control shRNA was subtracted from ΔCt value of circRNA targeting shRNA for corresponding circRNA/mRNA. Relative expression was measured by calculating $2^{-\Delta\Delta Ct}$.

3.7.4 Cell Death Assay for Colorectal Cancer Cell Lines

The methods are described for lentivirus from one cloned plasmid targeting a circRNA after the procedures explained in section 3.7.2. About 1 x 10^6 of puromycin resistant cells (HCT 116 and DLD-1) were seeded in a well of 6-well cell culture dish in 3 mL culture medium and cultured for 72 hours. Both floating and adhered cells were harvested together. The cells were stained with 7-AAD (BD Biosciences) and prepared for flow cytometry with CytoFLEX (Beckman Coulter) platform according to the manufacturer's protocol. CytoExpert Software package (that comes with the instrument) was used to set the experiment and get the percentage of dead cells in the circRNA targeted cell population. Cells, treated with 10 X Permeabilization Buffer (Thermo Fisher Scientific), was used as positive control for dead cells.

Table 3.8 qPCR primers for knockdown study of top essential circRNAs (IDT)

Hit ID	CircRNA/ mRNA	F.P./ R.P.	shRNA oligonucleotide sequences (5′-3′)	Amplicon size (bp)
Hit 1	CircRNA	F.P.	CCTGGAGGAGAGGATCGAGTT	125
		R.P.	TTCTGGATGGTCTGCTTGGA	
	mRNA	F.P.	TTTGGGAACGAGACTAGCCC	122
		R.P.	GGGAAGAAAATCTCCCGGCAT	
Hit 2	CircRNA	F.P.	GAGGCTGTGGACCGAGAGAT	187
		R.P.	TATCCAGCAGGTAACAGGGCT	
	mRNA	F.P.	TTTTCTGTTCTCGTCCGGGG	144
		R.P.	TGCTTTGGCACTGTTAGGGT	
Hit 4	CircRNA	F.P.	GACCCAACACAAATGGTTCCC	134
		R.P.	CTTGTCTGCAAACAGCTCCAA	
	mRNA		No mRNA	NA
Hit 5	CircRNA	F.P	GGAGGACTCCACACGCATT	126
		R.P.	TGAGGCAGGTACTTGGCATAC	
	mRNA	F.P.	CTGCGCCGTGTGAACCA	122
		R.P.	GAGCCCTTGGCAGCATTGAT	
GAPDH (control)	mRNA	F.P.	AAGGTGAAGGTCGGAGTCAAC	101
		R.P.	GGGGTCATTGATGGCAACAATA	

F.P.: Forward primer, R.P.: Reverse primer

The cell death for circRNA targeting shRNA was compared to the cell death for control shRNA.

3.7.5 Proliferation Assay for Colorectal Cancer Cell Lines

The methods are described for lentivirus from one cloned plasmid targeting a circRNA after the procedures explained in section 3.7.2. Puromycin resistant cells (HCT 116 and DLD-1) were seeded at two different numbers (10,000 and 25,000) in the wells of 12-well cell culture dish in 1 mL of culture medium. Cells were counted for four days in each 24 hours after seeding. Live cells were counted and the counts were normalized by the initial seeding number.

The live cell number for circRNA targeting shRNA was compared to the live cell number for control shRNA.

3.7.6 Colony Formation Assay for Colorectal Cancer Cell Lines

The methods are described for lentivirus from one cloned plasmid targeting a circRNA after the procedures explained in section 3.7.2. After puromycin selection, 200 cells (of HCT 116 and DLD-1) were seeded in a well of 12-well cell culture dish in 1 mL of culture medium. The cells were grown for seven days (medium was replaced if necessary). After day 7, media was removed and colonies were washed twice with 1 mL of PBS. About 400 µL of crystal violet (0.5%) was added to the colonies and culture dishes were incubated for 5 minutes. Extra crystal violet was removed by washing twice with 1 mL of PBS. Finally, colonies were counted and photographed. The colony number for circRNA targeting shRNA was compared to the colony number for control shRNA.

3.7.7 Tumorsphere Assay for Colorectal Cancer Cell Lines

The methods are described for lentivirus from one cloned plasmid targeting a circRNA after the procedures explained in section 3.7.2. After puromycin selection, 2,000 cells (HCT 116 and DLD-1) were seeded with 200 µL MammoCult™ Human Medium Kit (Stemcell Technologies) in a well of 96-well low attachment tissue culture dish. The tumorspheres were grown for eight days. The images of the tumorspheres were taken with ImageXpress®Micro (Molecular Devices) platform using MetaXpress 6 software package. The tumorspheres were harvested afterwards, cells from tumorspheres were trypsinized and the cell number was counted. The cell number for circRNA targeting shRNA was compared to the cell number for control shRNA.

3.7.8 Cell Death Assay for Other Cancer Types

Cell death assay for MDA-MB-231, MCF7, Du 145 and PC-3 cell lines were performed in the same way as described in section 3.7.4. The media compositions for the culture of mentioned cell lines are described in section 3.3.

3.7.9 Bioinformatics Study of Circular RNA Interaction with miRNA and Protein

The interaction study of essential circRNAs with miRNAs or proteins was conducted on 'CircInteractome' database. Default set up of 'miRNA Target Sites' option was used for circRNA-miRNA interactions and default set up of 'RBP on CircRNA' was used for circRNA-RBP interactions. The miRNAs and their target proteins were studied on 'TargetScan' (http://www.targetscan.org/vert_71/) database. Gene enrichment analysis was performed on miRNA interacting proteins using 'ToppGene' platform (https://toppgene.cchmc.org/).

4.0 Results

4.1 CircRNA shRNA Library Design, Construction, and Evaluation

To systematically identify essential circRNAs, we used lentivirus-based, pooled shRNA screening technology (Blakely *et al.*, 2011; Ketela *et al.* 2011; Marcotte *et al.*, 2012; Bassik *et al.*, 2013; Vizeacoumar *et al.*, 2013; Paul *et al.*, 2014; Paul *et al.*, 2016). Each unique short hairpin sequence in the shRNA library acts as a molecular barcode that can be sequenced to quantify the amount of each shRNA present after transducing cells with pooled lentiviral particles. A decrease in the frequency of specific shRNA sequences within the cell population can be used to identify essential circRNAs.

We designed a novel ~ 15K circRNA shRNA library that targets ~ 5K circRNAs. The library was designed by extracting human circRNA sequences from the 'circBase' database, which contains the most comprehensive collection of circRNAs compiled from several genomic studies as represented in Fig 4.1 (Glažar *et al.*, 2014). As of 2014, circBase contained 9,324 circRNAs and we aimed at developing shRNAs targeting all these circRNAs. Out of these

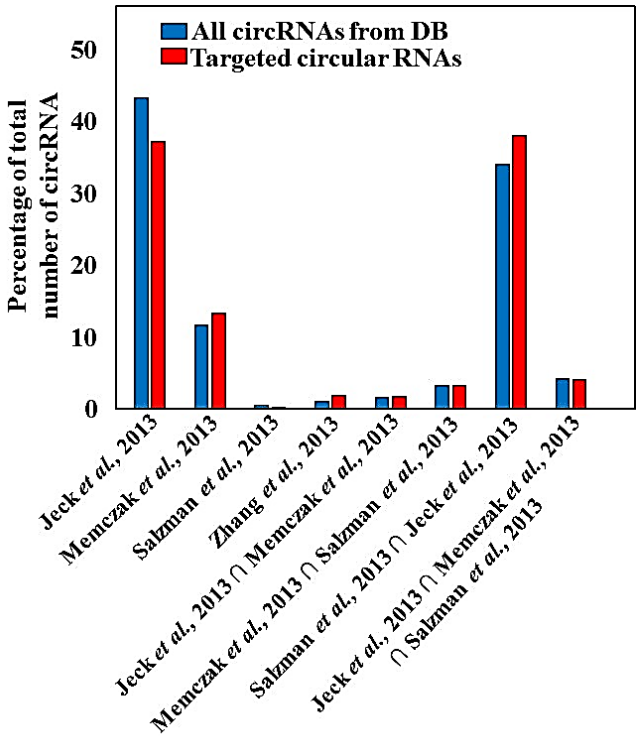

Figure 4.1 Different data sources of circRNAs. (A) Bar diagram showing number of circRNAs from different genomic studies collected from 'circBase'. Names of different genomic studies are plotted in X-axis and percentages of total number of circRNAs are plotted in Y-axis. The common circRNAs identified by multiple studies are represented with intersection (∩**)** sign.

circRNAs, the study of Jeck *et al.*, 2013 solely identified the highest 43.3% (4,038) of the circRNAs while 34.2% (3,184) of the circRNAs were overlapped between Salzman *et al.*, 2013 and Jeck *et al.*, 2013 studies (Fig. 4.1). Our targeted circRNAs had the highest fraction of 38.1% (1,893) from overlapping of Salzman *et al.*, 2013 and Jeck *et al.*, 2013 studies while 37.31% (1,855) from only Jeck *et al.*, 2013 study.

We used rules from the Broad Institute shRNA design to create a circRNA shRNA library for all 9,324 circRNAs. For this, we designed 21-mers of respective shRNAs targeting

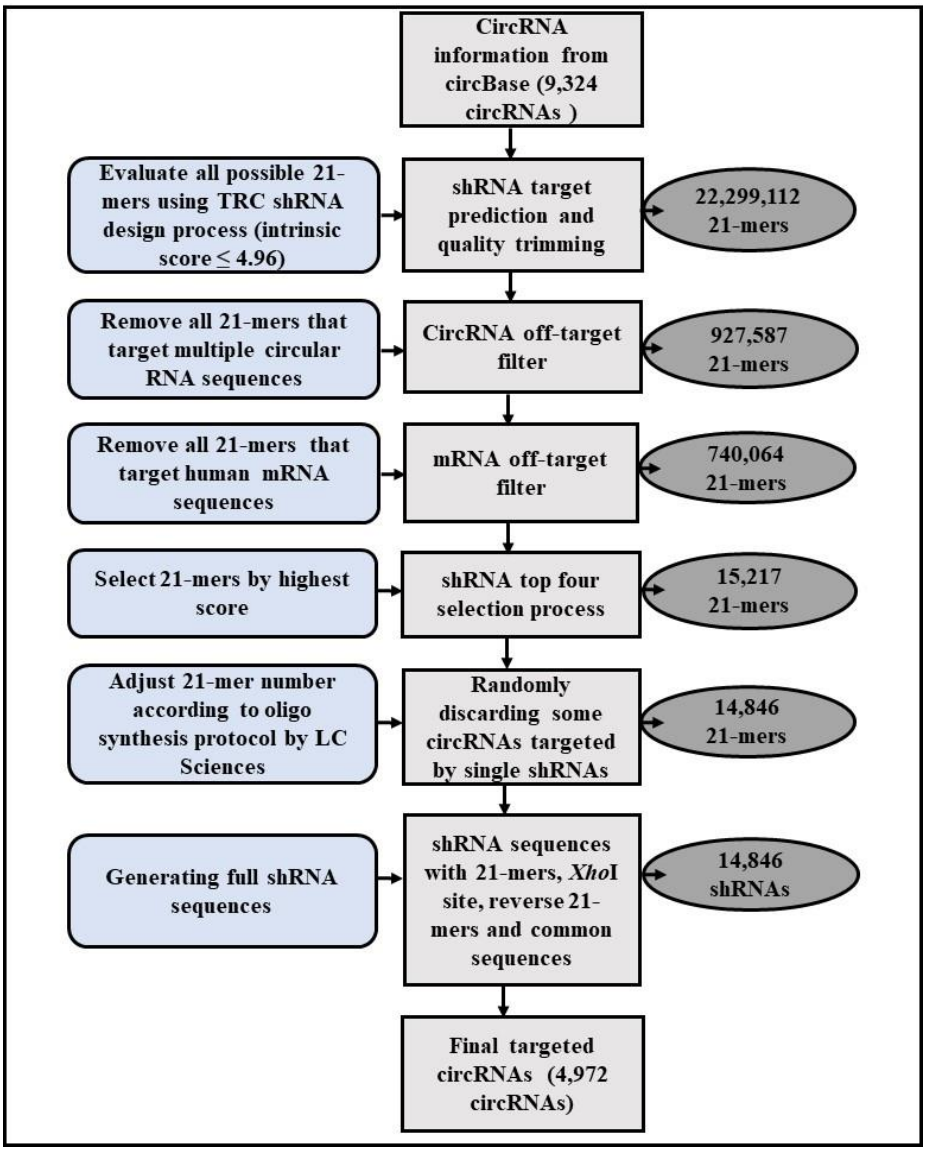

Figure 4.2 Design of shRNA library to target circRNAs. Flow chart showing a pipeline to design shRNA library to target circRNAs. The rationales and respective procedures for all filtering and/or selection steps to design the circRNA shRNA library are described. The resulting shRNA number after each filtering and/or selections step is presented.

the circRNAs that resulted in total of 22,299,112 21-mers (Fig. 4.2). Our circRNA off-target filter, mRNA off-target filter, shRNA top four selection process and the oligo synthesis protocol of LC Sciences (restriction in chip coverage) reduced this number of shRNAs to 927,587, 740,064, 15,217 and 14,846 21-mers, respectively (Fig. 4.2). The circRNA off-target filter ensured that each shRNA in the circRNA shRNA library targeted only one circRNA. We also designed the circRNA shRNA library to specifically target human circRNAs and not human mRNAs. The mRNA off-target filter ensured that no shRNA in the circRNA shRNA library targeted any human mRNAs. This filtering helped to select the shRNAs targeting the back-spliced region alone (Fig. 4.3A). Based on their intrinsic scores, only the top four sequences were retained for a given circRNA for which have more than four shRNAs (Fig. 4.2). The filtering steps also resulted in different numbers of shRNAs per circRNA ranging from one to four. Finally, we designed 14,846 shRNAs for 4,972 circRNAs. Our library did not cover 4,352 circRNAs from the circRNA dataset (Fig. 4.2).

We constructed the circRNA shRNAs using these filters, resulting in 82% of circRNAs being targeted by at least two shRNAs (Fig. 4.3B). Most circRNAs were targeted by four

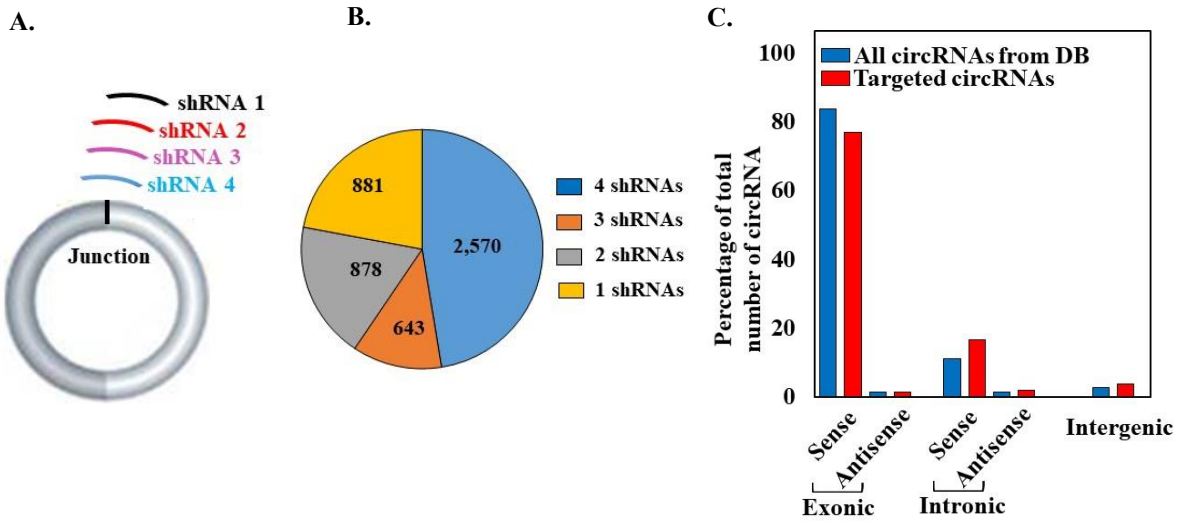

Figure 4.3 Key features of circRNA shRNA library. **(A)** Illustration showing the design of shRNA library to specifically target unique junctions of circRNAs. **(B)** Pie chart showing comparison of circRNAs targeted by different numbers of shRNAs in the designed circRNA shRNA library. **(C)** Bar diagram showing the classification and respective fractions of circRNAs based on their genomic origins for circRNAs in both 'circBase' dataset and targeted circRNA set. Different genomic locations are plotted in X-axis and percentages of total number of circRNAs are plotted in Y-axis.

shRNAs (51.7%) in the library and a small number of circRNAs were only targeted by one shRNA (17.7%) (Fig. 4.3B). The majority of targeted circRNAs were from genic origin (96.4%) in comparison to a smaller fraction from intergenic region (3.6%) (Fig. 4.3C). Moreover, both E- and I- circRNAs from genic circRNA population, originated mostly from sense strand of the gene. The fraction of both sense and antisense I-circRNAs and intergenic circRNAs were increased in targeted circRNAs in comparison to respective fractions in circBase dataset (Fig. 4.3C). The targeted E-circRNA (both sense and antisense) fraction was reduced in comparison to circBase E-circRNA fraction. Our circRNA shRNA library covered circRNAs present on multiple chromosomes (Fig. 4.4A). All chromosomes had a very similar distribution of circRNAs in both circBase circRNAs and final targeted ones. Fractions of circRNAs from chromosome 1 and chromosome 2 were comparatively high, while those circRNAs from chromosome X and chromosome Y were lower (Fig. 4.4A). Our library also targeted circRNAs of different sizes from very small (~ 50 bp) to very large ones (~ 1 mbp) as shown in Fig. 4.4B.

We synthesized the shRNA oligonucleotides for the designed circRNA shRNA library with the facilities of LC Sciences. We PCR-amplified the shRNA library oligonucleotides and cloned amplicons into the TRC cloning plasmid p.LKO.1 using the highly efficient 'NEBuilder' method. The 142 bp shRNA library amplicon and the ~ 7 kb plasmid fragment were analyzed on an agarose gel (Fig. 4.5A, B). During construction of library, we ensured sufficient representation of the library oligonucleotides. We validated our cloned circRNA shRNA library by restriction digestion with *Xho*I. Expected fragments from the restriction digestion (~ 6.5 kb, 310 bp and 190 bp) were detected using an agarose gel (Fig. 4.6A). We also validated our library using NGS method. We PCR-amplified the circRNA shRNA library from the plasmid, checked the amplicon size on an agarose gel (205 bp), extracted the library amplicon, and sequenced the amplicon using the Ion Torrent platform (Fig. 4.6B). We obtained total 245,884 shRNA library sequence counts, which represented ~ 17-fold coverage of the library. Approximately 92% (13,750) of our designed shRNA sequences were present in the raw sequence reads, which covered 97.4% (4,842) of targeted circRNAs (Fig. 4.7).

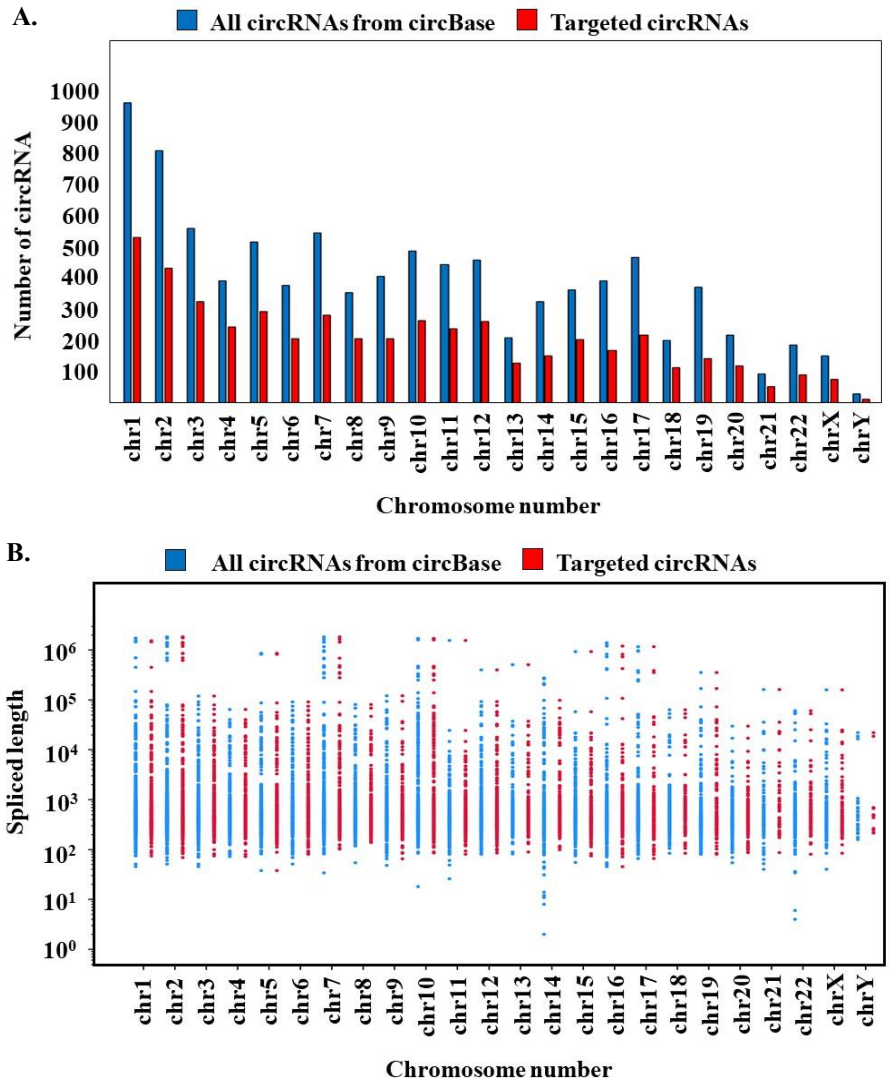

Figure 4.4 Position and size distribution of circRNAs. (A) Bar diagram showing distribution of circRNA numbers over different chromosomes. Chromosome numbers are plotted in X-axis and numbers of circRNAs are plotted in Y-axis. **(B)** Dot plot showing the size distribution of circRNAs over different chromosomes. In both cases, circRNAs targeted by shRNA library were compared to circRNAs from 'circBase' dataset. Chromosome numbers are plotted in X-axis and spliced lengths of circRNAs are plotted in Y-axis.

4.2 Analysis of Library Sequencing Methods

We used the circRNA shRNA library to identify essential circRNAs in a colorectal cancer cell line (HCT 116). We chose this cell line because the Vizeacoumar lab previously performed similar pooled shRNA screens using these cells (Vizeacoumar *et al.*, 2013). Briefly, HCT 116 cells were transduced with a MOI (0.3 to 0.4) under conditions where each cell was transduced by only one lentivirus and the library was covered ~ 600-fold in the screen. The

initial cell numbers for transduction were calculated based on MOI. Cells were cultured with puromycin to remove un-transduced cells. After selecting transduced cells, genomic DNA was prepared from the circRNA shRNA containing a cell population isolated at different time points during the screening procedure. CircRNA shRNA library sequences were PCR-amplified from the genomic DNA using plasmid-backbone directed universal primers (Blakely *et al.*, 2011; Ketela *et al.* 2011). The abundance of each circRNA shRNA was quantified using Ion Torrent sequencing method and compared among an initial, a medium and the last time point of the screening process. The following results on optimization of shRNA library sequencing methods are described from previous works by Islam *et al.*, 2017.

Sequencing of the library plasmid showed that ~ 30.1% of sequences in the raw sequence reads were prematurely terminated despite the circRNA shRNA library amplicon being purified

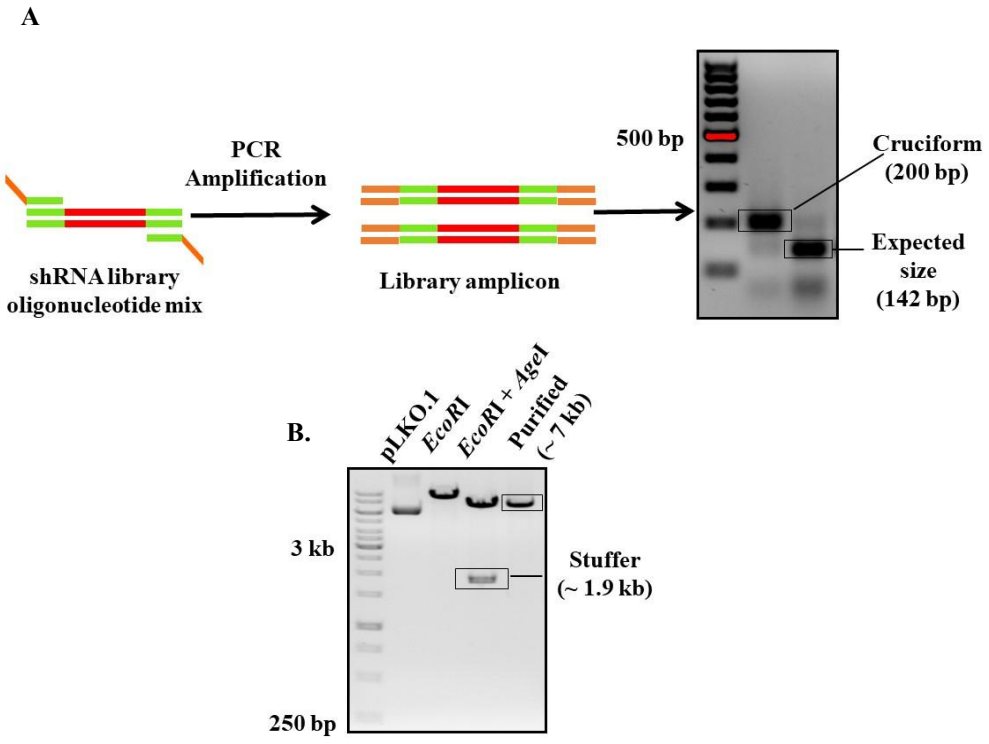

Figure 4.5 Preparation of circRNA shRNA library. **(A)** Schematic and gel image showing preparation method of oligonucleotides of circRNA shRNA library. The oligonucleotides, synthesized by LC Sciences, were PCR-amplified to attach Gibson Assembly linkers and detected as 142 bp product using an agarose gel. The library amplicon was extracted and used to clone into pLKO.1 plasmid. **(B)** Gel image showing different products by digestion of pLKO.1 plasmid with *EcoR*I and *Age*I. The double-digestion released stuffer fragment (~ 1.9 kb) from the plasmid that helped to clone the library amplicon into the remaining ~ 7 kb product (detected and extracted from agarose gel) using NEBuilder method.

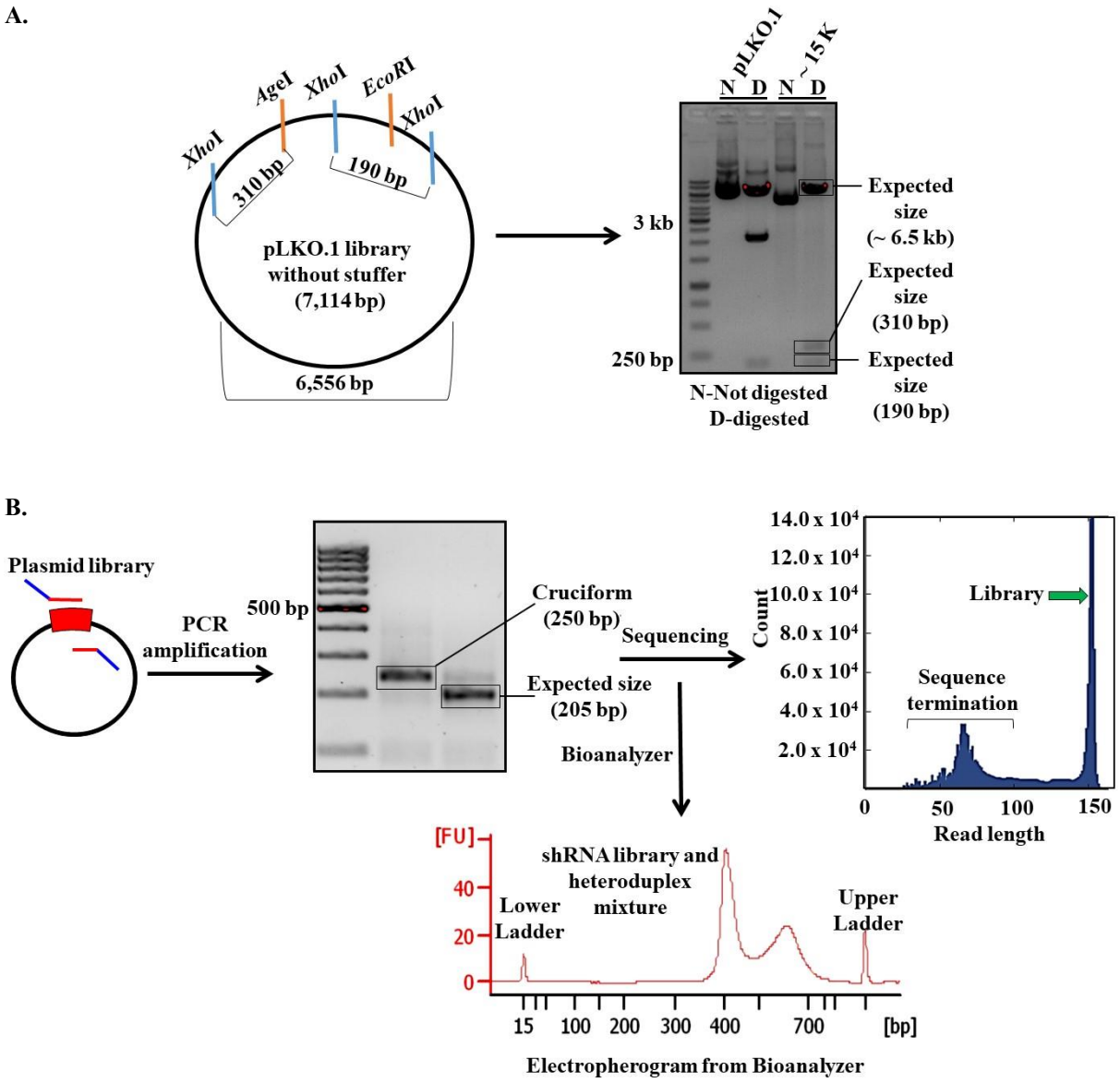

Figure 4.6 Validation of constructed circRNA shRNA library. **(A)** Schematic and gel image showing validation method of circRNA shRNA library with *Xho*I digestion. The digested products were compared between pLKO.1 plasmid and plasmid library using an agarose gel. **(B)** Representative schematic, agarose gel, electropherogram from Bioanalyzer and read-length histogram showing validation method of circRNA shRNA library by Ion Torrent sequencing. For the read-length histogram, sequence length is plotted in X-axis and counts of library sequences is plotted in Y-axis. The length of the shRNA library is 152 bp, excluding the A and P1 sequences.

as a single band using an agarose gel (Fig. 4.6B). We also observed 13% polyclonal reads and 23% low quality reads. In case of sequencing from genomic DNA sample, the read length histogram of the circRNA shRNA library sequencing showed that only 21.2% of the sequence

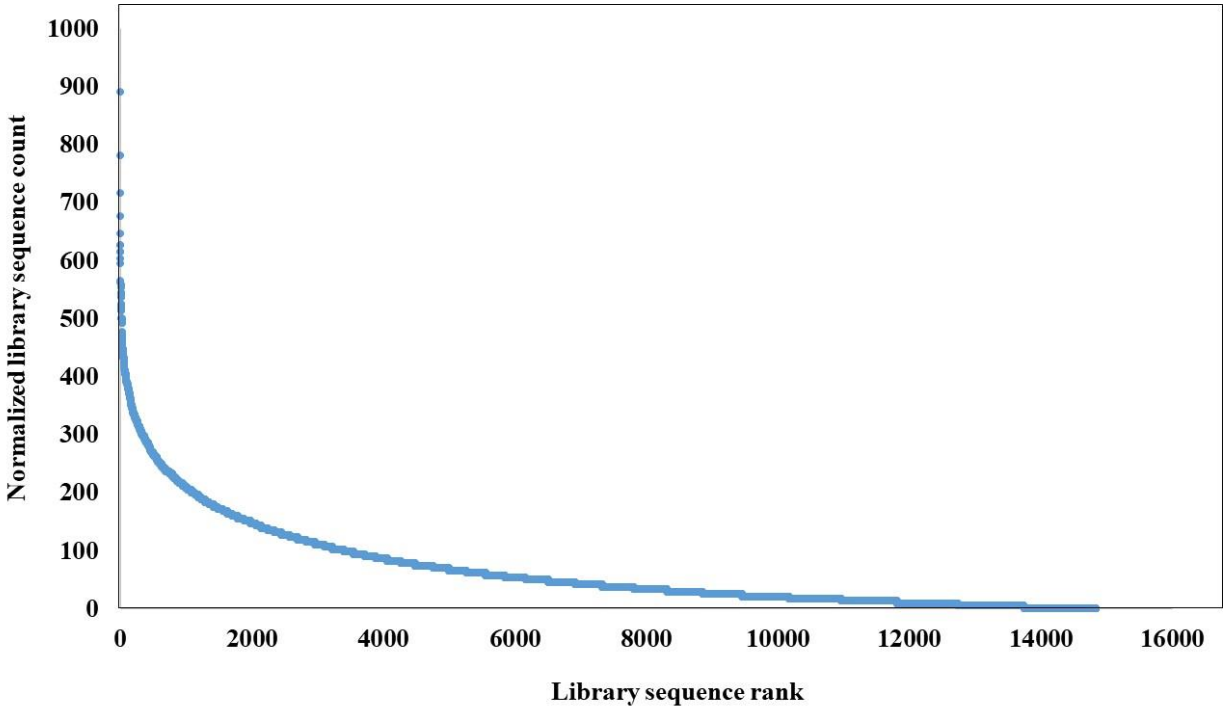

Figure 4.7 Library sequence representation in NGS data of library plasmid. Line chart showing counts of library sequences from Ion Torrent sequencing of circRNA shRNA plasmid library. Each of the library sequence count was normalized to per million of total library sequence count. Ranks of the library sequences are plotted in X-axis and normalized library sequence counts are plotted in Y-axis.

output resulted in usable reads (Fig. 4.8A). Of this, only 84% of our ~ 15K circRNA shRNA library sequences could be detected. This represented only ~ 10-fold coverage of the library and was not sufficient for statistical analyses. Analysis of incomplete reads revealed that termination of these sequences started at around 30 bp, which corresponded to the beginning of the hairpin loop (Fig. 4.8A). This suggested that formation of hairpin structures was responsible for this premature sequence termination. A second major peak of terminated sequences also occurred around 65 bps (Fig. 4.8A). Variation in circRNA shRNA sequence began at 57 bp and sequence terminations at this point likely resulted from the low-quality scores associated with misidentified polyclonal Ion Sphere Particles (ISPs). We characterized the library amplicon using capillary microfluidic gel electrophoresis by running samples on Bioanalyzer and observed multiple DNA fragments (Fig. 4.6B and Fig. 4.8A). We hypothesized that secondary structures and/or mixed heteroduplex templates were responsible for these results. The Ion PGM^{TM} System uses the following steps to sequence DNA: emulsion

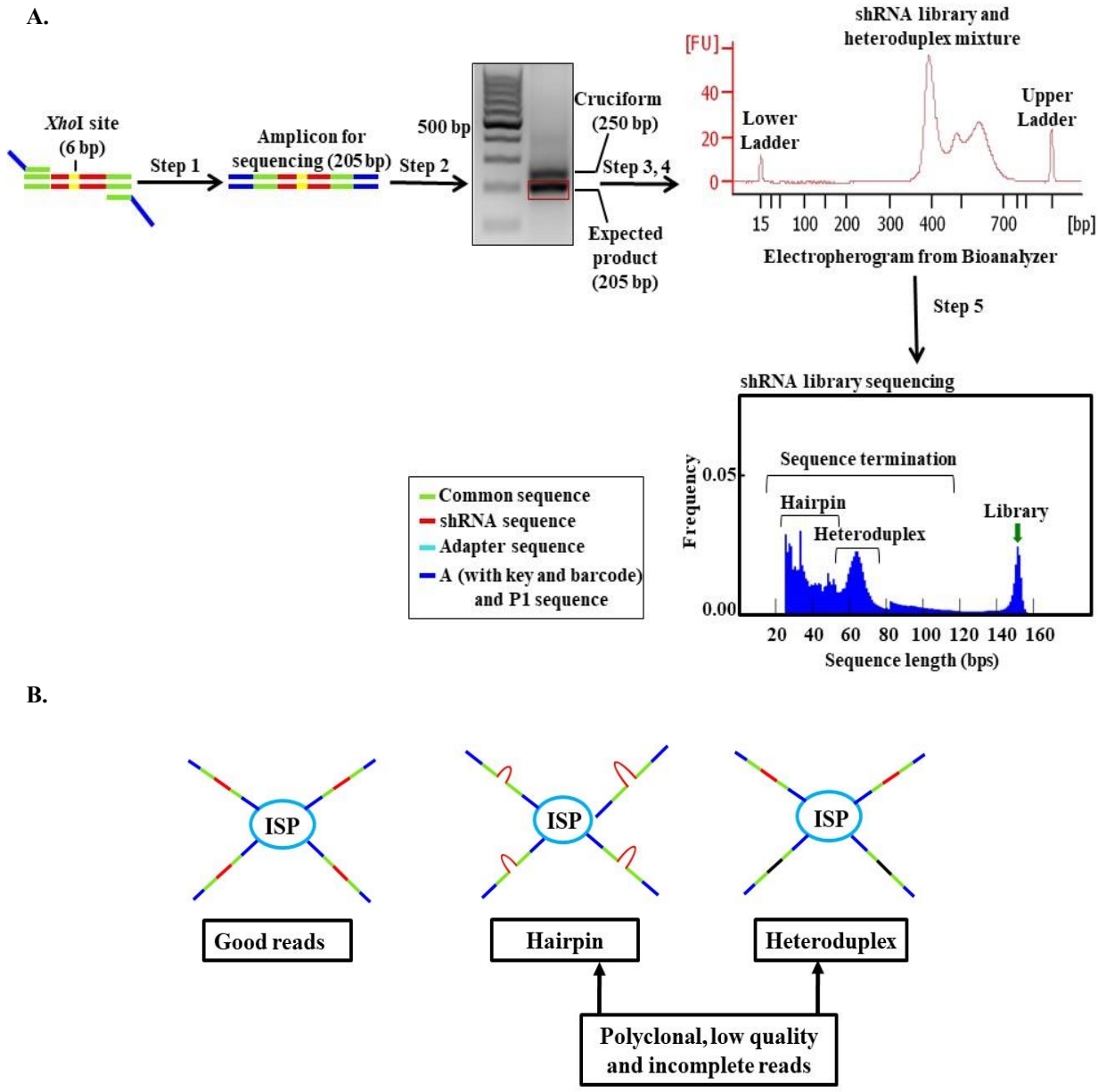

Figure 4.8 Steps involved in shRNA library sequencing. **(A)** Workflow showing steps involved in sequencing shRNA library from the genomic DNA. Step1: PCR amplification of shRNA library from gDNA. Step 2: PCR purification of the amplified library. Step 3: Gel extraction of the 205 bp library amplicon. Step 4: Quality assessment of the library using a Bioanalyzer. Step 5: shRNA library sequencing Ion Torrent platform. Representative agarose gel, electropherogram from Bioanalyzer and read-length histogram from Ion Torrent sequencing are shown. Sequence lengths are plotted in X-axis and respective frequencies are plotted in Y-axis. The length of the shRNA library is 152 bp, excluding the A and P1 sequences. **(B)** Illustrations of ideal and non-ideal ISP conditions in Ion Torrent sequencing. Non-ideal situations were assumed to reduce sequencing quality. Adapted from Islam *et al.*, 2017.

PCR, enrichment of template positive ISPs, and NGS. In principle, the clonal amplification step should have resulted in ISPs that were coated with multiple copies of a single sequence from the library (Fig. 4.8B). During sequencing, all copies of the template on an ISPs were expected to be single-stranded and anneal to the sequencing primer. But at the time of primer annealing, single-stranded template had formed a hairpin structure on some ISPs, which resulted in termination of the polymerization reaction and shortened the anticipated sequencing reads (Fig. 4.8B). Additionally, this hairpin structure caused de-synchronization between clonal fragments attached to a single ISP during the incorporation of new nucleotides in the synthesizing strands. Inappropriate nucleotide insertion at one or several flows led to either polyclonal or low-quality reads. Moreover, heteroduplex formation caused by the annealing of two different circRNA shRNA sequences resulted in ISPs being coated by two different templates (Fig. 4.8B). This resulted in polyclonality and generated many polyclonal and low-quality sequencing reads. As a consequence, incomplete, low quality, and polyclonal reads were found in the final sequencing output.

Overall, our initial sequencing of the circRNA shRNA library indicated that the sequencing was hindered by hairpin structure and heteroduplex formation. Sequencing of this library resulted in a high proportion of prematurely terminated, polyclonal and low-quality sequencing reads. This is a major shortcoming when sequencing is carried out using medium-throughput sequencers such as Ion PGM™ System. In an effort to reduce this problem, we developed strategies to eliminate the hairpin structure and reduce heteroduplex formation.

To eliminate the hairpin structure, we cleaved the circRNA shRNA sequence in half by taking advantage of the *Xho*I restriction site present in the hairpin loop region (Moffat, J. *et al.*, 2006). In order to PCR-amplify these half-shRNA sequences, we designed a *Sal*I adapter and ligated this to the half-shRNA to generate an adapter-ligated half-shRNA product (Fig. 4.9). During the ligation of *Sal*I adapter to *Xho*I-digested product, self-ligated products were also formed between *Xho*I-digested products or *Sal*I adapters (Fig. 4.9). The smaller *Sal*I self-ligated product was removed during the PCR purification step (Fig. 4.9). To eliminate the *Xho*I self-ligated product, we digested the self-ligated *Xho*I products with *Xho*I and removed digested products using PCR purification and gel extraction (Fig. 4.9). *Xho*I (C/TCGAG) and *Sal*I (G/TCGAC) recognition sites are compatible to each other and when ligated they create a unique site (CTCGAC), which cannot be recognized by *Xho*I and/or *Sal*I restriction enzymes.

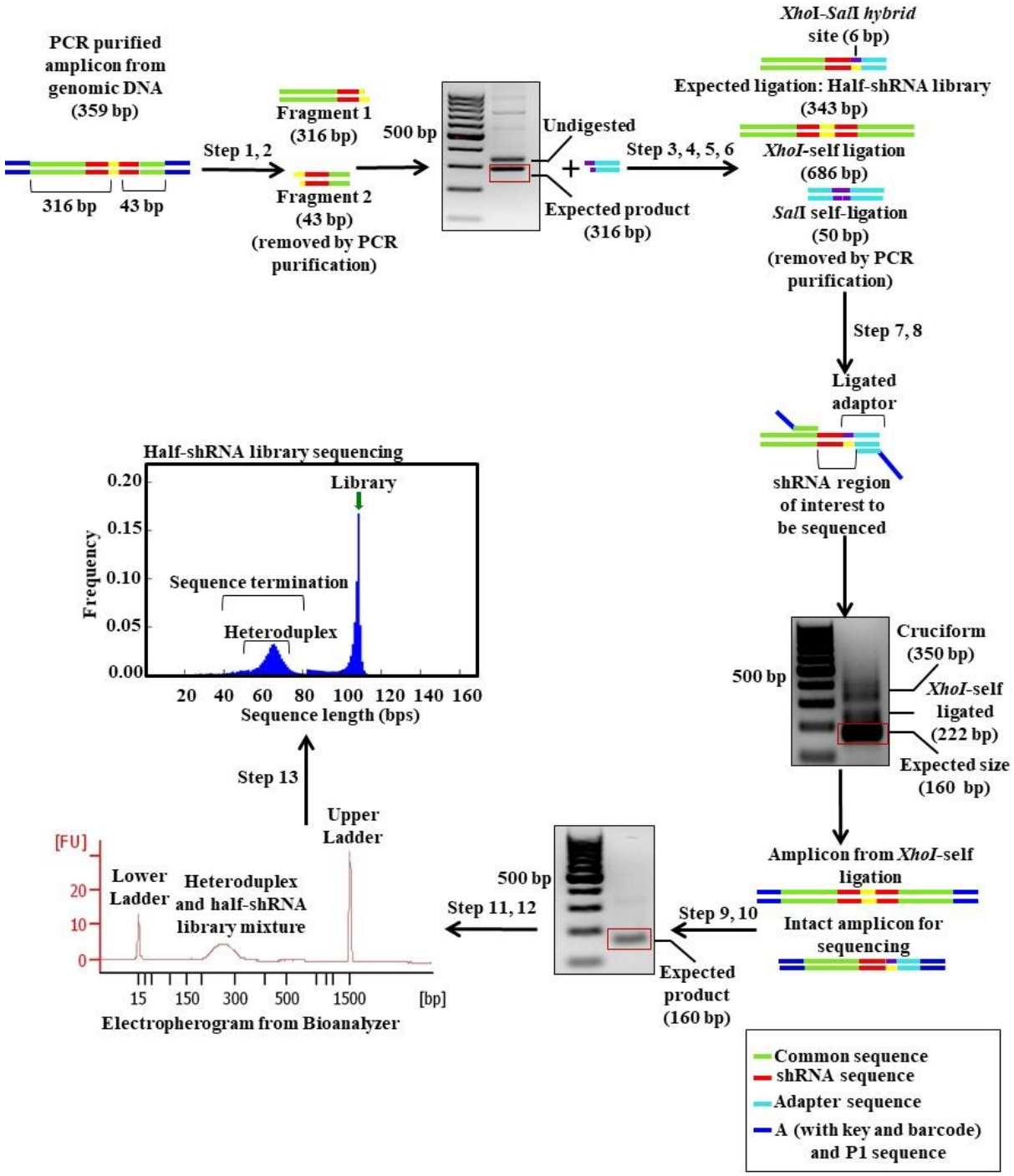

Figure 4.9 Elimination of the hairpin in shRNA library. Workflow showing steps to eliminate hairpin in the shRNA library. Step 1: *Xho*I digestion of purified PCR amplicon from gDNA. Step 2: PCR purification of *Xho*I-digested product. Step 3: Gel extraction of expected 316 bp *Xho*I-digested amplicon. Step 4: Preparation of the *Sal*I adapter. Step 5: Adapter ligation to *Xho*I-digested amplicon. Step 6: PCR amplification of the ligated product. Step 7: PCR barcode labeling of the ligated product and detection of 160 bp PCR product on an agarose gel. Step 8: PCR purification of the barcoded PCR product. Step 9: *Xho*I digestion of

the amplicon from *Xho*I self-ligated product. Step 10: PCR purification after digestion of the amplicon from *Xho*I self-ligation. Step 11: Gel extraction of the 160 bp barcoded product. Step 12: Quality assessment of the barcoded product on Bioanalyzer. Step 13: Half-shRNA library sequencing on Ion Torrent platform. Representative agarose gel, electropherogram from the Bioanalyzer and read-length histogram from Ion Torrent sequencing are shown. Sequence lengths are plotted in X-axis and respective frequencies is plotted in Y-axis. The length of the half-shRNA library is 108 bp excluding A and P1 sequences. Adapted from Islam *et al.*, 2017.

This new site makes it possible to remove only the *Xho*I self-ligated product but not the expected product. By digesting the shRNA in half, we eliminated premature sequence stalling at around 30 bp as shown by the loss of the peak in the read length histogram (Fig. 4.9). The half-shRNA sequencing increased the number of reads of the expected library and 45.9% of the sequence output was usable reads compared to 21.8% in shRNA sequencing. These sequences covered 94.7% of the ~ 15K circRNA shRNA library with a 26-fold coverage. The adapter-ligated half-shRNA product did not generate a sharp clean peak when analyzed using the capillary microfluidic gel electrophoresis on Bioanalyzer (Fig. 4.9). We assumed that the broad sized peak was due to heteroduplex formation. Consistent with the Bioanalyzer data, we found that a second major peak of terminated reads around 65 base pairs still remained, suggesting the existence of heteroduplexes (Fig. 4.9). To reduce heteroduplex formation, we decreased the number of PCR cycles from 28 to 15 cycles, a strategy suggested for decreasing heteroduplex formation (Ruano and Kidd, 1992; Thompson *et al.*, 2002; Meyer and Kircher, 2010; Rentero Rebollo *et al.*, 2014; Gorbacheva, *et al.*, 2015). The reduction in PCR cycles decreased the yield of the amplicon, and thus we used magnetic bead-based purification (Agencourt AMPure XP) to minimize DNA loss (Fig. 4.10). This solid phase reverse immobilization technique was advantageous for low concentration DNA clean up and we successfully obtained a 95% yield after purification. This modification eliminated the heteroduplex formation which resulted in a pure preparation of the library amplicon as seen in the Bioanalyzer analysis (Fig. 4.10). Consistent with this, we found that the peak at 65 bp was reduced (Fig. 4.10). From the read length histogram we observed that the PCR optimization substantially removed the formation of heteroduplex and increased the expected library reads (Fig. 4.10). In case of improved half-shRNA sequencing, 82.5% of the sequence output was usable reads as compared to 45.9% in just half-shRNA sequencing or 21.8% in shRNA sequencing (Fig. 4.10). These sequences

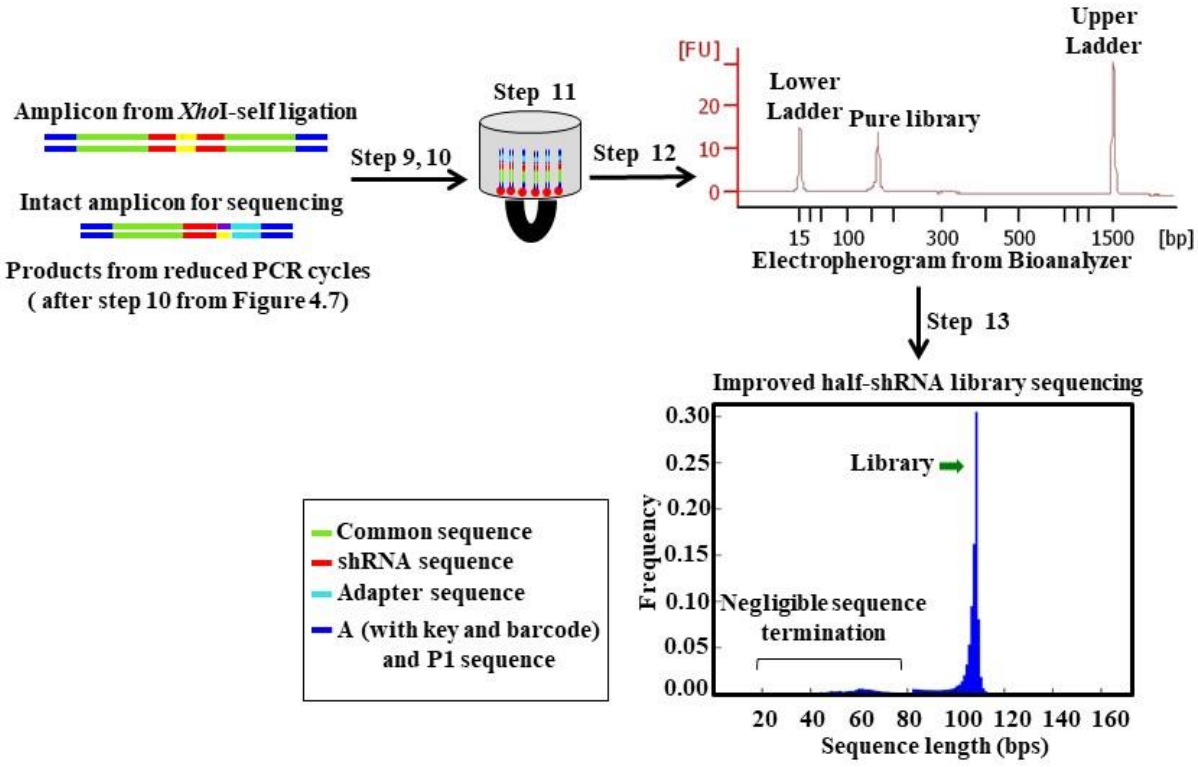

Figure 4.10 Reducing heteroduplex formation in half-shRNA library. Workflow showing steps to reduce heteroduplex formation in half-shRNA library. Step 1-10: same as in Fig. 4.9. In step 8, reduced number of PCR cycles were used which was the exception from the previous procedure. Step 11: Magnetic bead-based purification of the barcoded product. Step 13: Quality assessment of the barcoded product on the Bioanalyzer. Step 14: Half-shRNA library sequencing on Ion Torrent platform. Representative agarose gel, the electropherogram from Bioanalyzer, read-length histogram from Ion Torrent sequencing are shown. Sequence lengths are plotted in X-axis and respective frequencies are plotted in Y-axis. The length of the half-shRNA library is 108 bp excluding A and P1 sequences. Adapted from Islam *et al.*, 2017.

covered 98.7% of our ~ 15K shRNA library with a 55-fold coverage. Overall, these results suggested that a considerable amount of non-usable reads could be eliminated by removing hairpin and heteroduplex formation. To show that optimized PCR conditions eliminated library heteroduplex formation, we also used a pooled Genome-scale CRISPR Knock Out (GeCKO) library (Shalem *et al.*, 2014). As these CRISPR libraries consist of mixed-oligonucleotides, they also suffer from the formation of heteroduplexes structures during PCR amplification (Fig. 4.11A-C). These are much larger libraries and might not be compatible with Ion PGM™ System due to limited throughput. Still we expected that elimination of heteroduplex structures would improve the multiplexing capabilities in medium throughput instruments such as

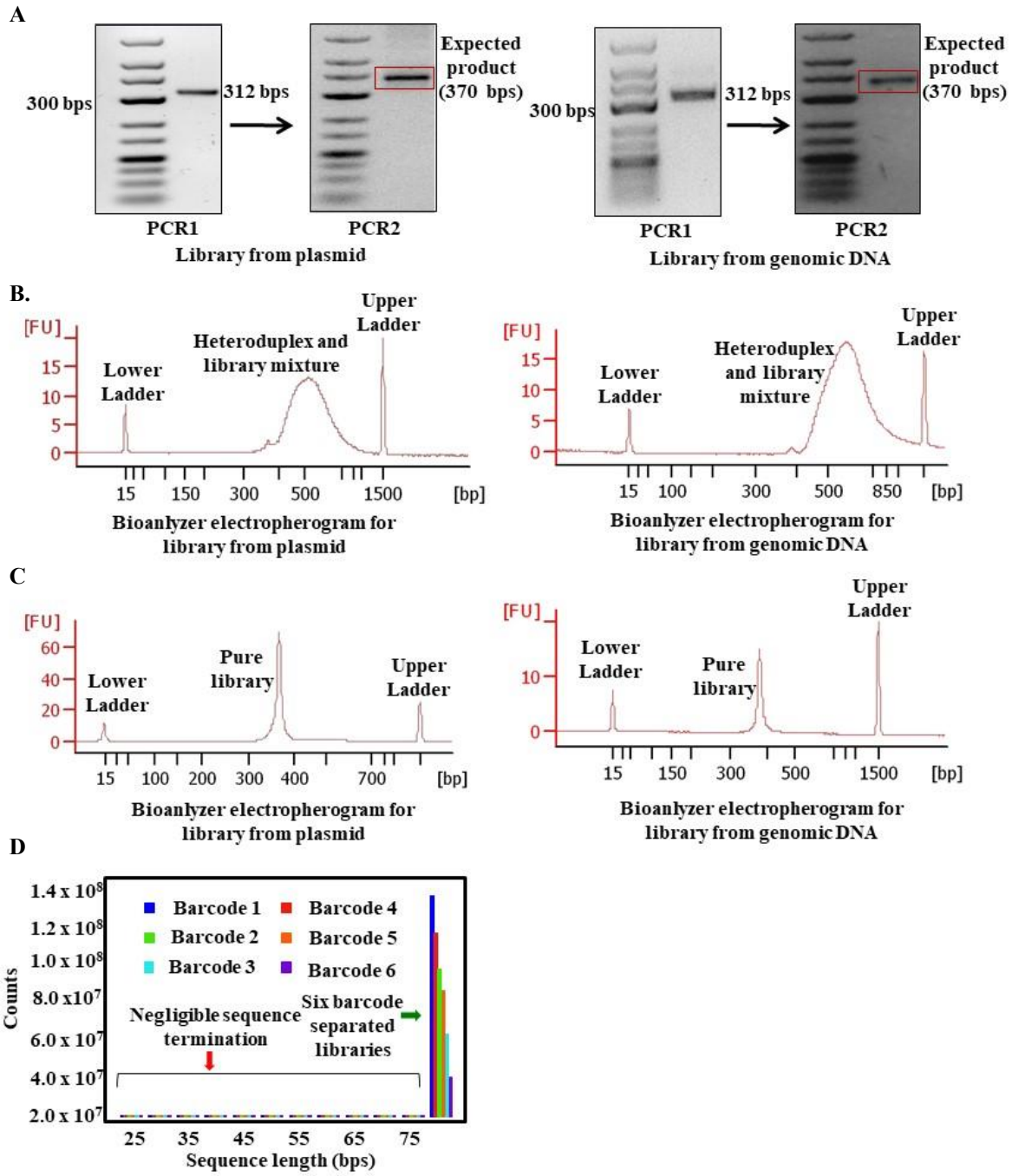

Figure 4.11 Steps involved in NGS of GeCKO library. (A) Gel image of barcoded GeCKO library preparation from plasmid or gDNA. In both cases, 312 bb product from PCR1 was detected on an agarose gel and 370 bp product from PCR2 was extracted from the gel. **(B)** Bioanalyzer electropherogram showing formation of heteroduplex in barcoded CRISPR library from both plasmid and gDNA. **(C)** Bioanalyzer electropherogram showing elimination of heteroduplex formation in GeCKO library from both plasmid and gDNA by reduced PCR cycles. **(D)** Read-length histogram showing Illumina sequencing results for the GeCKO library amplified from plasmid. The length of the GeCKO library is 76 bp after barcode separation.

Sequence lengths are plotted in X-axis and sequence counts are plotted in Y-axis. Six different barcodes represent six different dilutions of the library, where consecutive dilutions caused comparatively lower sequence counts. Adapted from Islam *et al.*, 2017.

Illumina NextSeq500. Therefore, we extended our method as a validation strategy for quality assessment of the GeCKO library prior to Illumina sequencing. In addition, a detailed protocol in preparing CRISPR libraries for Illumina sequencing will also benefit researchers as such resources are still limited. Sequencing of the GeCKO library requires two steps of PCR reactions. Reduction of cycles for both PCR reactions to amplify the library from plasmid or genomic DNA, decreased the ~ 520 bp DNA smear and resulted in an increase of the ~ 370 bp product (Fig. 4.11B, C). Sequencing of the GeCKO library from plasmid on the Illumina platform by adapting our procedure showed minimal amounts of incomplete reads, indicating that our method was equally applicable for larger libraries as well (Fig. 4.11D).

Overall, with the improvement of the methods, it was clearly observed that polyclonal, low quality, and terminated reads were gradually decreased, while the intended library reads were increased (Fig. 4.12A). The library-fold coverage was also increased though there was no comprehensive difference among identified library sequences by these methods (Fig. 4.12B, C).

4.3 Analysis of Pooled Screening Results to Identify Essential Circular RNAs

We PCR-amplified circRNA shRNA library for eight different samples (two replicates of initial time point T0, three replicates of middle time point T1 and three replicates of last time point T2). Library amplicons from these time points were sequenced using the Ion Torrent sequencing method. Overall, we obtained ~ 2,978,646 usable reads with ~ 70 fold coverage of the circRNA shRNA library. This represents 99% of the shRNA library that we designed to target ~ 4,972 circRNAs. Different counts of circRNA shRNA library sequencing are summarized in Table 4.1. For further analysis, the raw count of each sequence of the library at a time point was normalized to per million of total library sequence count, which was then used to calculate DCC scores. We focused our analysis on sequences that decreased over time (dropouts) to identify essential circRNAs in HCT 116 cells. We found 8,471 library sequences whose counts decreased to different extents (positive dropout score range: 711.65 to 0.008463) from T0 to T2 (Fig. 4.13 and Fig. 4.14A). Interestingly, 6,376 library sequences were enriched to different extents (negative dropout score range: -0.51694 to -542.575) from T0 to T2 (Fig.

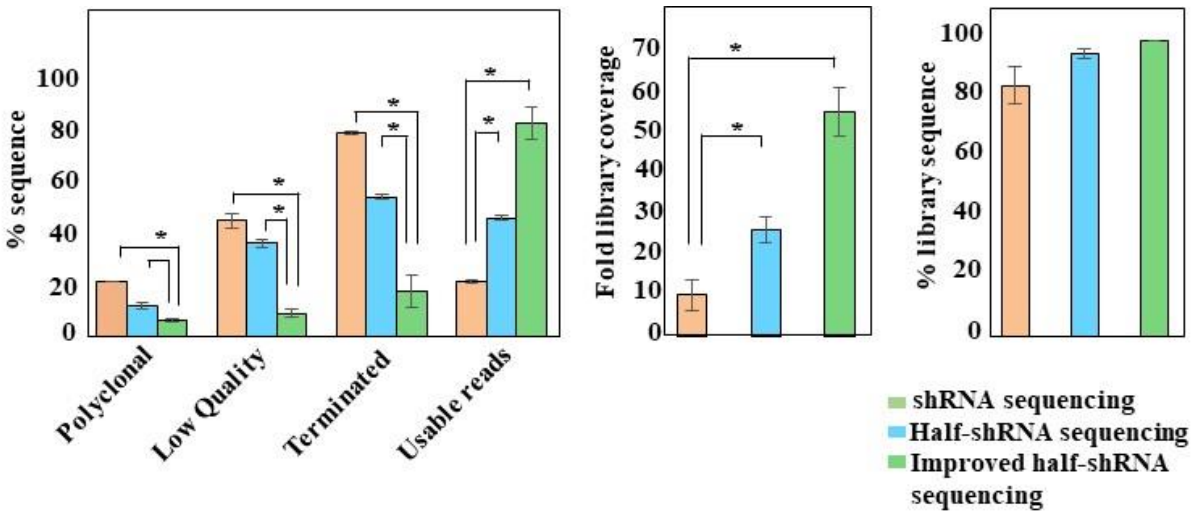

Figure 4.12 Comparison of the quality of Ion Torrent sequencing across different methods. (A) Bar diagram showing comparison of metrics from shRNA, half-shRNA and improved half-shRNA library sequencing. Different types of sequence reads are plotted in X-axis and percentages of sequence reads are plotted in Y-axis. Polyclonal reads are presented as percentage of initial loading and enrichment. Low quality reads are presented as percentage of total sequence output after loading, enrichment and initial polyclonal removal. Terminated and barcoded sequence reads are presented as percentage of total filtered reads by Ion Torrent server. **(B)** Bar diagram showing comparison of library fold coverage from shRNA, half-shRNA and improved half-shRNA library sequencing. Different types of sequencing methods are plotted in X-axis and respective fold coverages of the library are plotted in Y-axis. **(C)** Bar diagram showing comparison of identified library sequences from shRNA, half-shRNA and improved half-shRNA library sequencing. Different types of sequencing methods are plotted in X-axis and respective percentages of identified library sequences (in raw sequence reads) are plotted in Y-axis. The experiments were done in duplicates, means ± standard deviation (error bars). *$P < 0.05$ (Student's t-test). Adapted from Islam *et al.*, 2017.

4.13 and Fig. 4.14B). However, with a stringent bootstrapping of DCC scores and considering at least two shRNAs of the similar effect per circRNA, we identified 107 essential circRNAs (bootstrap p-value ≤ 0.05) whose loss-of-function caused lethality to HCT 116 cells. Normalized library sequence counts at different time points are shown for top five essential circRNAs in Fig. 14.15 with their annotations presented in Table 4.2.

Out of 107 essential circRNAs, 79.4% (76.6% sense 2.8% antisense) were exonic, 14.9% were intronic (14% sense and 0.9% antisense) and 5.6% were intergenic (Fig. 4.16A and Fig. 4.16B). The percentage of essential circRNAs that we identified correlated with the abundance of different categories (exonic, intronic and intergenic) of targeted circRNAs. As functions of these circRNAs are not known, we decided to study parental genes of genic circRNAs (94.3%

Table 4.1 Different counts of circRNA shRNA library sequencing are

Library parameters	**Counts**
Total library sequence count	2,979,646
Fold coverage	66.9
Percentage of the total library	98 %
Percentage of targeted circRNA	99 %

of top 107). Interestingly, 25.17% of genes were involved in signal transduction (Fig. 4.16C), 25% were related to cell communication (Fig. 4.16D), and 21.7% did not have any known association with a biological processes (Fig. 4.16D). Furthermore, many of these genes had expression in both cancer and normal tissues with an enrichment in colon (57.6%), rectum (45.7%), and colorectal cancer (47.8%) (Fig. 4.16E). Out of the 107 essential circRNAs, 21 circRNA had bootstrap p-value ≤ 0.01. Some of the parental genes of these 21 circRNAs had differential expression between normal and cancer tissues (analyzed using patient-derived expression data from 'TCGA') (Fig. 4.16F). For example, *PMM1* and *WBSCR22* represent parental genes of circRNA hsa_circ_0005703 and hsa_circ_0005588 respectively. *PMM1* is downregulated in most cancers (including colon cancer) but *WBSCR22* is upregulated in most cancers (including colon cancer) (Fig. 4.17). This indicated possibilities that essential E-circRNAs could utilize mechanisms independent of the expression of their parental genes.

4.4 Results of the Cloning of shRNA Oligonucleotides into the shRNA Expression Plasmid

To validate the circRNA shRNA screen without bias, we took the top five essential circRNAs and cloned individual shRNA oligonucleotides into pLKO.1 plasmid to test them on a one-on-one basis (Fig 4.18). These five circRNAs represent three E-circRNAs, one I-circRNA, and one intergenic circRNA. We successfully cloned two shRNA oligonucleotides for four of the top five essential circRNAs. The shRNA oligonucleotides for the I-circRNA (hsa_circ_0092290) could not be cloned for unknown reasons. For the cloning, we PCR-amplified the shRNA oligonucleotides to attach Gibson Assembly linkers that were detected as a 142 bp product using an agarose gel. The 142 bp shRNA oligonucleotide amplicons were cloned into ~ 7 kb *Xho*I- and *Age*I-digested pLKO.1 plasmid. The successful cloning was confirmed by performing colony PCR and checking the PCR products on agarose gel, checking

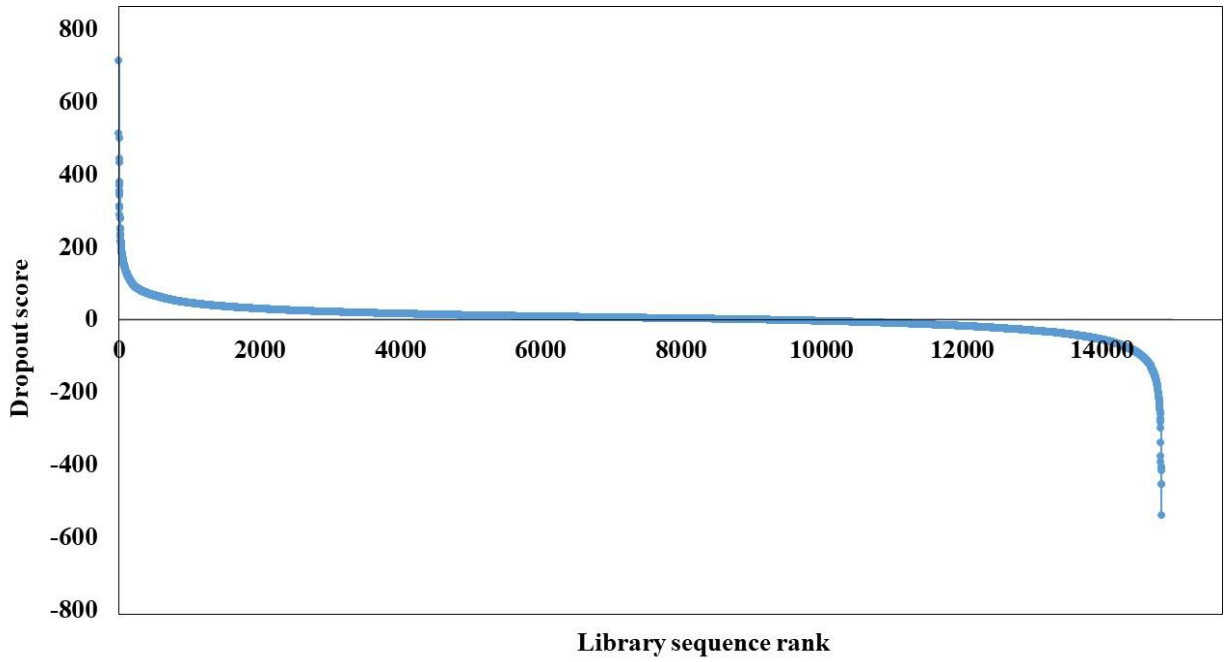

Figure 4.13 Library sequence dropout and enrichment identification from NGS data. Line chart showing dropout scores of library sequences from sequencing of circRNA shRNA library from genomic DNA. Each of the library sequence count was smoothed by five (five is added to the count) and normalized to per million of total sequence count. Differences of the normalized counts between T0 and T2 time points were calculated for each of the library sequence and ranked accordingly. Ranks of the library sequences are plotted in X-axis and difference of the normalized counts are plotted in Y-axis. Positive scores represent sequence dropouts and negative scores represent enrichment of sequences.

cloned plasmids on agarose gel, and Sanger sequencing of cloned insert of shRNA sequences. In total, we cloned eight oligonucleotides of circRNA shRNAs into pLKO.1 which targeted three E-circRNAs and one intergenic circRNA.

4.5 Results of CircRNA shRNA Assays

HCT 116 cells were transduced with the relevant circRNA shRNA expressing lentivirus followed by puromycin selection. The extracted total RNA from transduced HCT 116 cells (by both control and treatment shRNAs) showed 28S and 18S ribosomal RNA bands, indicating intact RNA sample quality. Small RNAs and some degraded RNA appeared as a lower molecular weight smear (Fig. 4.19A). To specifically amplify circRNAs, we designed divergent primers, which amplify circRNAs and not corresponding mRNAs. To test these primers, we did reverse transcription PCR (RT-PCR) of total RNA samples and analyzed

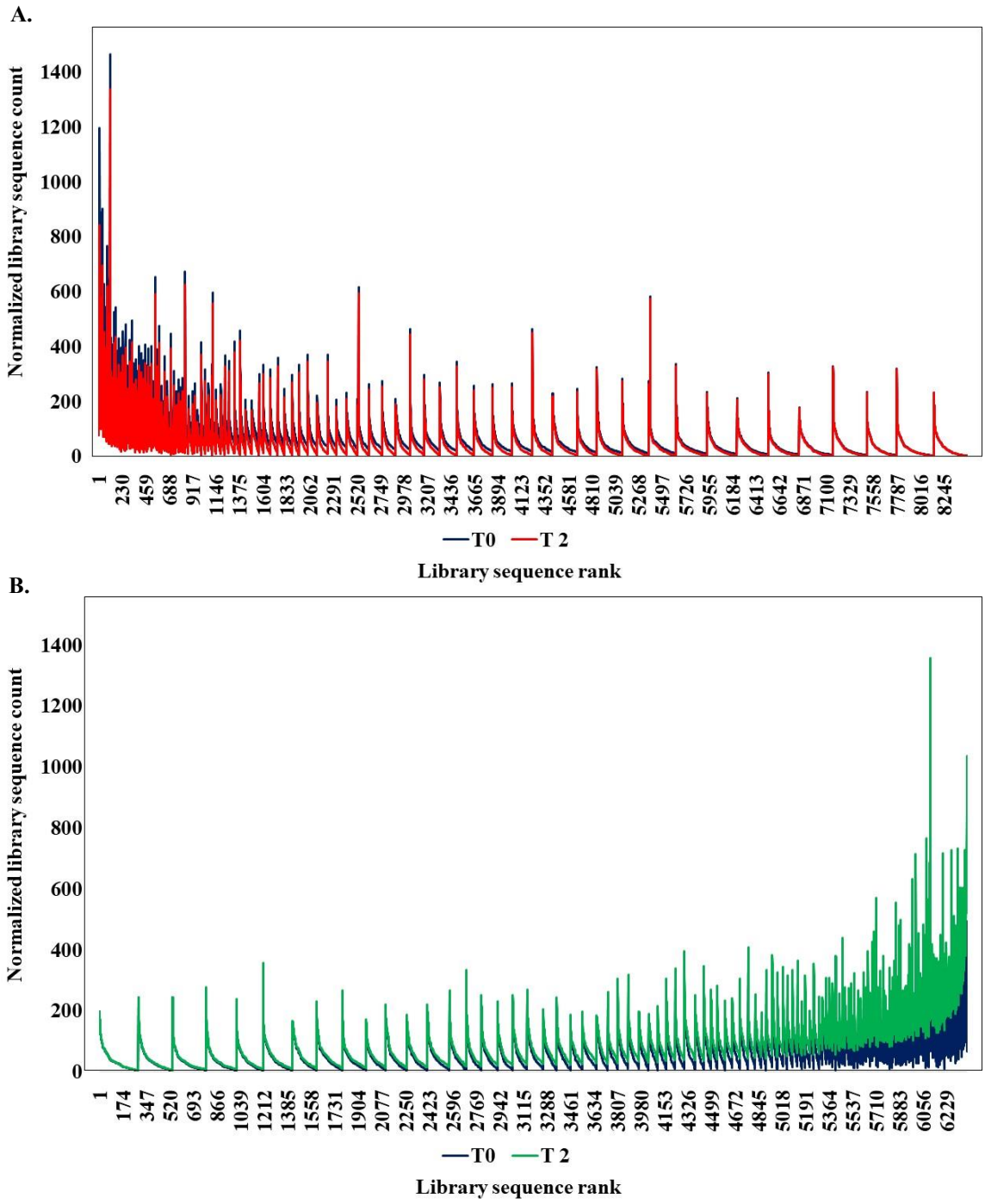

Figure 4.14 Comparison of library sequence abundance between initial and last time points of screening procedure. **(A)** Bar diagram showing comparison between T0 and T2 time points based on normalized library sequence counts to find out sequence dropouts. Ranks of the library sequences are plotted in X-axis and normalized library sequence counts are plotted in Y-axis. **(B)** Bar diagram showing comparison between T0 and T2 time points based on normalized library sequence counts to find out library sequence enrichments. Ranks of the library sequences are plotted in X-axis and normalized library sequence counts are plotted in Y-axis.

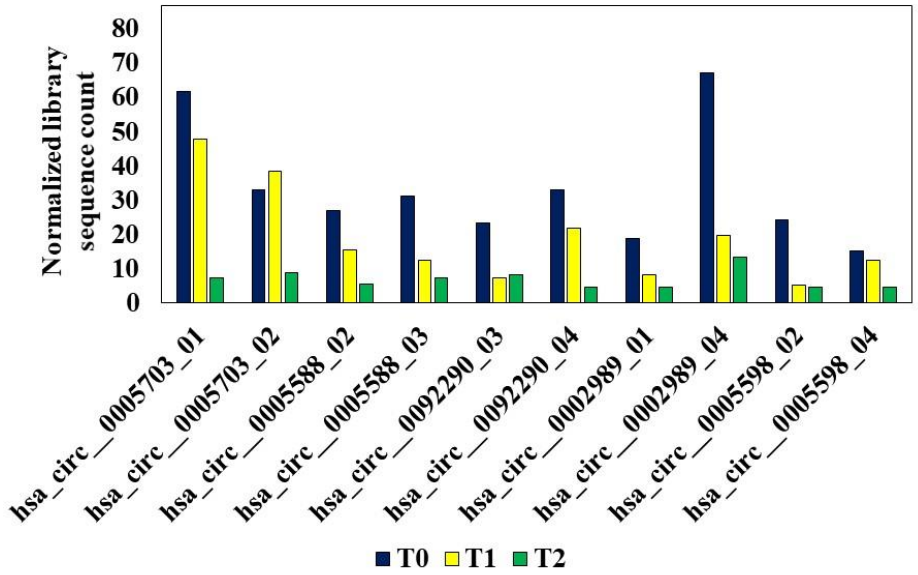

Figure 4.15 Identification of library sequence dropouts. Bar diagram showing the library sequence counts of T0, T1 and T2 time points for the top five essential circRNAs. Each of the library sequence count was smoothed by five (five is added to the counts) and normalized to per million of total sequence count. Names of the two shRNA sequences per circRNA are plotted in X-axis and respective counts are plotted in Y-axis.

Table 4.2 List of top five essential circRNAs from screening of HCT 116 cell line

Screen rank	**Bootstrap *p*-value (five decimal digits)**	**CircRNA ID**	**Gene**	**Spliced length**	**Origin type**
1	0.00000	hsa_circ_0005703	*PMM1*	269	Exonic (single exon)
2	0.00042	hsa_circ_0005588	*WBSCR22*	276	Exonic (single exon)
3	0.00108	hsa_circ_0092290	*SCRIB*	300	Intronic
4	0.00113	hsa_circ_0002989	No gene	358	Intergenic
5	0.00182	hsa_circ_0005598	*RPS5*	339	Exonic (single exon)

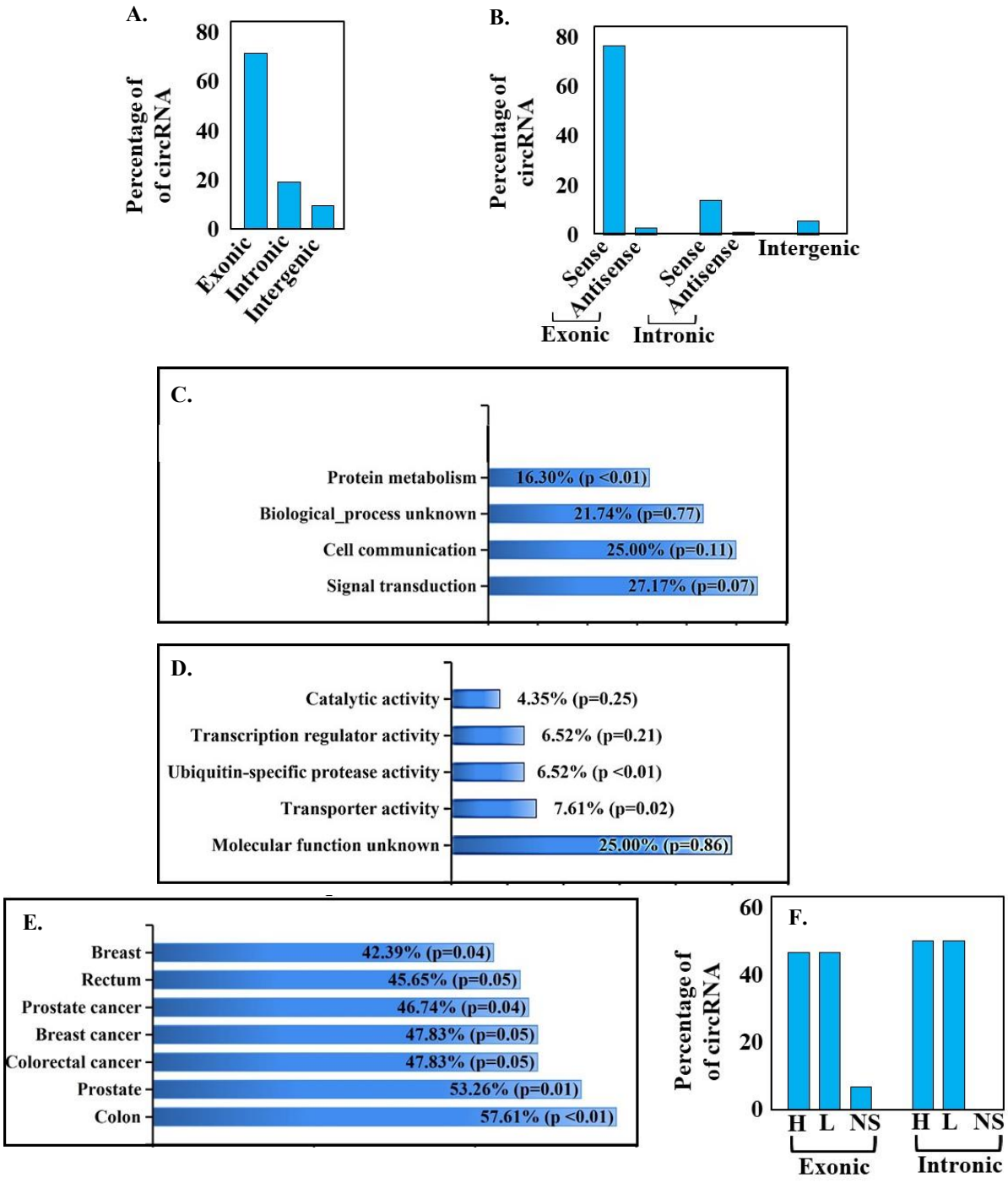

Figure 4.16 Features of top circRNA hits. (A) Bar diagrams showing fractions of circRNAs based on their genomic origins for top 107 circRNA hits and for top 21 circRNA hits. For both cases, different genomic locations are plotted in X-axis and percentages of total circRNAs (107 or 21 circRNAs respectively) are plotted in Y-axis. Bar diagrams showing enrichment of parental genes of E-circRNAs from top 107 circRNA hits based on **(C)** molecular function **(D)** biological process and **(E)** site of expression. Percentages of the enriched genes are plotted in X-axis and respective molecular functions, biological processes and sites of expressions are

plotted in Y-axis. The corresponding *p*-values were calculated by FunRich software. **(F)** Bar diagram showing fractions of parental genes of E-circRNAs from top 21 circRNA hits based on differential expression between normal and cancer tissues (colon adenocarcinoma) provided by 'TCGA'. Types of expression in X-axis and percentages of total circRNAs (21 circRNAs) are plotted in Y-axis. H: higher expression in cancer tissues than normal tissues, L: lower expression in cancer tissues than normal tissues, NS: no significant expression difference between normal and cancer tissues.

amplicon sizes. We also checked product sizes for corresponding mRNAs (for E-circRNAs) by using mRNA specific primers (convergent primers). For example, hsa_circ_0005588 had amplicons of 187 bp and 147 bp for two different primer sets and corresponding mRNA (*WBSCR22* gene) had amplicons of 144 bp and 124 bp for two different primer sets (Fig.

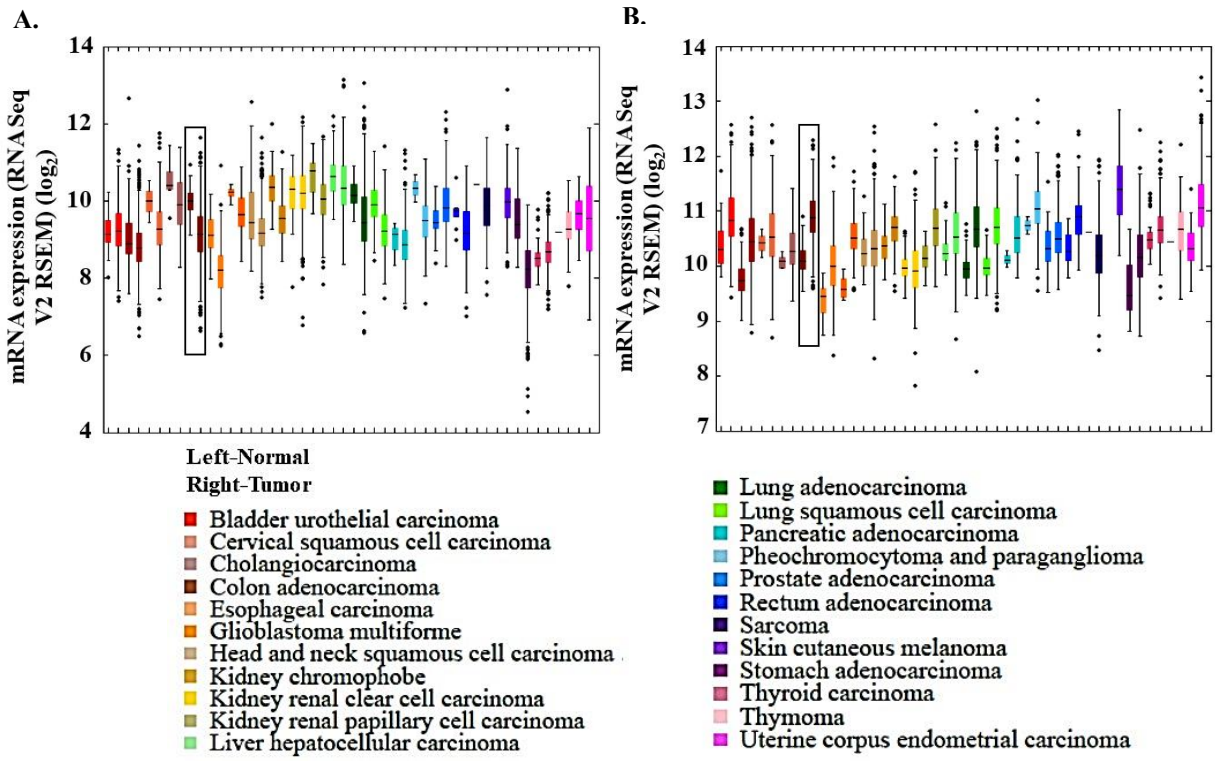

Figure 4.17 Expression study of parental genes of top essential E-circRNAs in multiple malignancies. (A) Box-and-whisker plot showing *PMM1* (parental gene of has_circ_0005703) expression in different cancer types and matching normal tissue controls provided by 'TCGA'. **(B)** Box-and-whisker plot showing *WBSCR22* (parental gene of has_circ_0005588) expression in different cancer types and matching normal tissue controls provided by 'TCGA'. The normal and tumor conditions are plotted in X-axis and expression data are plotted in Y-axis. The expression data provided by 'TCGA' showed mRNA expression levels, which had been analyzed using RNA Seq V2 and was normalized using RSEM normalization according to 'TCGA' standards.

4.19A). The RT-PCR did not show any primer dimer, indicating that the amplicons were useful for qPCR analysis. The qPCR amplification plots showed single types of amplicons, indicating the primer sets specifically amplified the E-circRNAs or the corresponding mRNAs. Additionally, absence of primer dimer or any other product was confirmed from melt curves (Fig. 4.19B).

Interestingly, we found that E-circRNAs had higher Ct values than related mRNAs, indicating lower expression of circRNAs than linear counterparts in un-transduced HCT 116 cells (Fig. 4.20A). We also confirmed that circRNA shRNAs specifically targeted only the E-

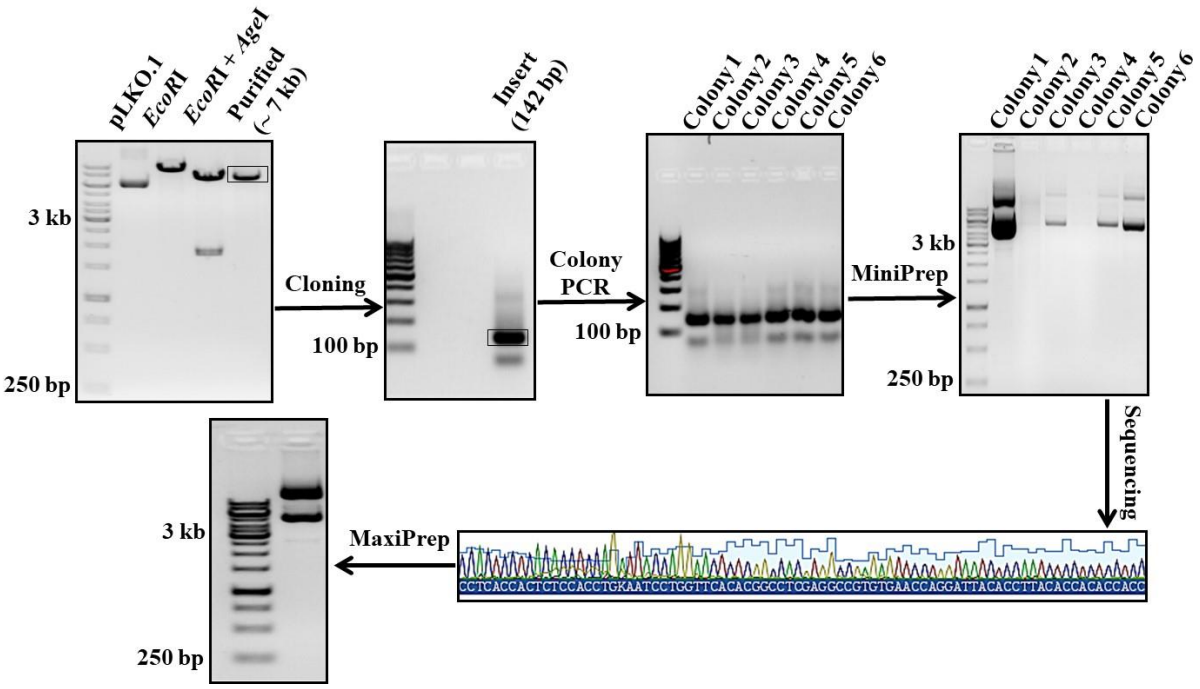

Figure 4.18 Cloning of shRNA oligonucleotides targeting top essential circRNAs. Gel images and sequencing chromatogram showing different steps of cloning of an shRNA oligonucleotide into pLKO.1 plasmid. The oligonucleotide was PCR-amplified to attach Gibson Assembly linkers and detected as 142 bp product using agarose gel. The double-digestion (with *EcoR*I and *Age*I) of pLKO.1 plasmid released stuffer fragment (~ 1.9 kb) from the plasmid that helped to clone the oligonucleotide amplicon into the remaining ~ 7 kb product (detected and extracted from an agarose gel) using Gibson Assembly method. Bacterial colonies that were successfully transformed by the cloned plasmids, were picked from bacterial plates and used for colony PCR to check the size of the shRNA oligonucleotide insert amplicon (142 bp). After size confirmation, the cloned plasmids were produce by MiniPrep method and detected on agarose gel as ~ 7 kb product. The cloned insert for shRNAs sequence was further confirmed by Sanger sequencing as represented by the chromatogram. The appropriate colonies were used to produce the cloned plasmids in large scale by MaxiPrep method, which were detected on agarose gel as ~ 7 kb product.

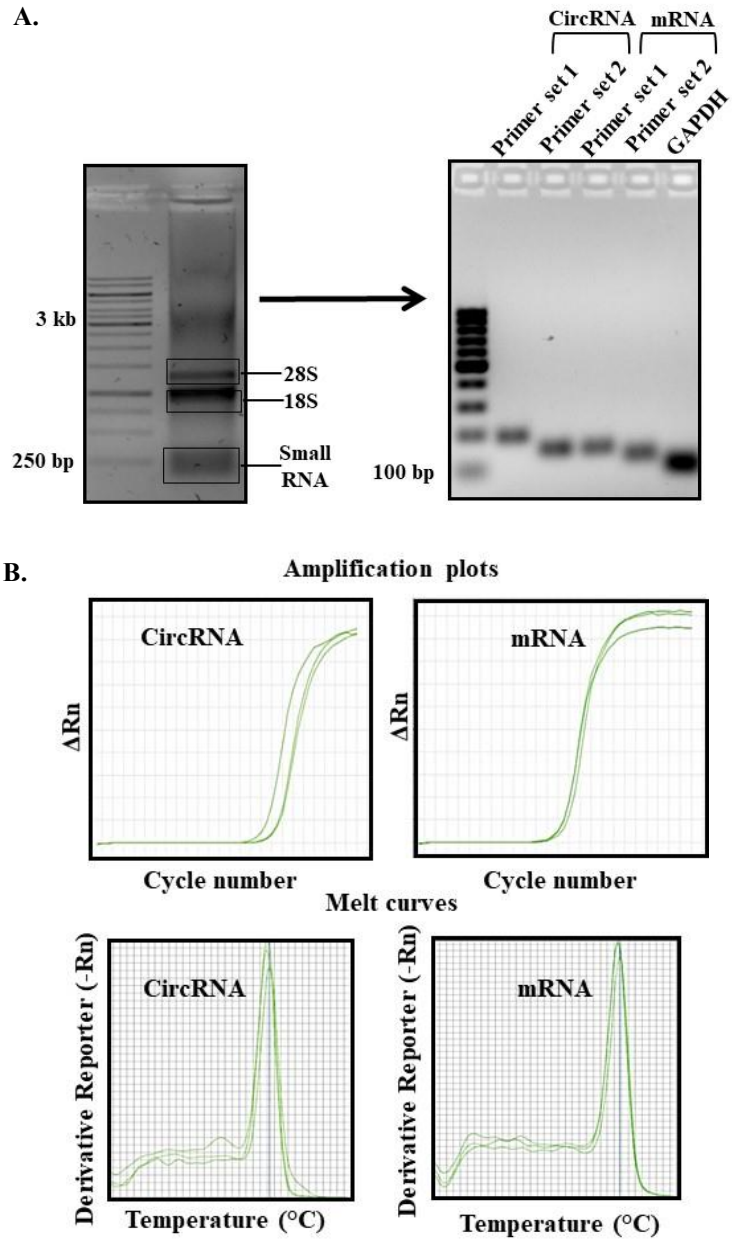

Figure 4.19 Expression study of a top essential circRNAs and corresponding mRNAs of top essential E-circRNAs. **(A)** Gel images showing quality assessment of total RNA extraction from HCT 116 cells and PCR amplicons from reverse transcription PCR (RT-PCR) of an E-circRNA and respective mRNA by using specific primers. **(B)** Amplification plots and melt curves showing specificity and purity of the amplicon of the circRNA and mRNA in qPCR study. In the amplification plot, cycle numbers are plotted in X-axis, and ΔRn (normalized SYBR Green fluorescence, Rn) values are plotted in Y-axis. In the melt curves, temperatures are plotted in X-axis and Derivative Reporter (-Rn) values are plotted in Y-axis. Amplification plot and melt curve are presented for three technical replicates from the sample.

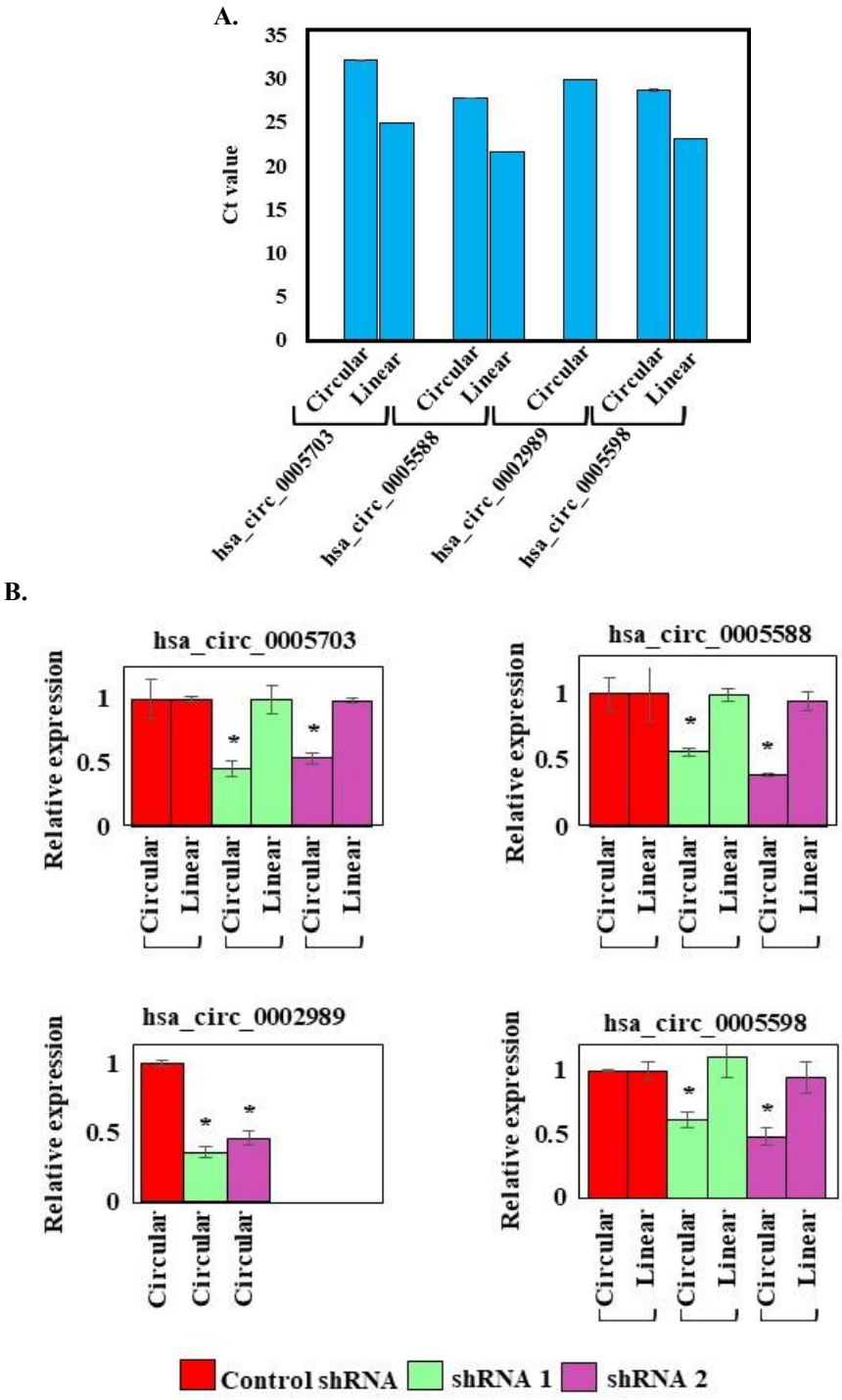

Figure 4.20 Evaluation of knockdown of top essential circRNAs and corresponding mRNAs of top essential E-circRNAs using divergent and convergent primers. (A) Bar diagram showing expression of circRNAs and corresponding linear mRNAs (for E-circRNAs) in HCT 116 cells. The circRNA and mRNA names are plotted in X-axis and critical threshold (Ct) values are plotted in Y-axis. **(B)** Bar diagrams showing qPCR results of relative

expression of essential circRNAs and their corresponding linear mRNAs (for E-circRNAs). Knockdown and matching control cells were seeded and cultured for 72 hours before qPCR. Note that hsa_circ_0002989 does not have any linear mRNA as it comes from intergenic region. The experiments were done in triplicates, means ± standard deviation (error bars). *$P < 0.05$ (Student's t-test).

circRNAs and not the corresponding linear mRNAs (Fig. 4.20B). We observed that shRNA caused significant reduction in expression level of all four circRNA hits (hsa_circ_0005703, hsa_circ_0005588, hsa_circ_0005598, hsa_circ_0002989). In contrast, mRNAs (for hsa_circ_0005703, hsa_circ_0005588, hsa_circ_0002989) did not show any significant change in expression.

4.6 Results of Cell Death Assay in Colorectal Cancer Cell Lines

To examine whether the shRNA-mediated knockdown of circRNAs caused any effect on cell viability, we used two colorectal cancer cell lines (HCT 116 and DLD-1). After lentiviral transduction and puromycin selection, we showed using 7-AAD staining and flow cytometry that shRNAs for E-circRNAs (hsa_circ_0005703, hsa_circ_0005588, hsa_circ_0005598) caused a significant increase in cell death in comparison to the control shRNA in HCT 116 cell line (Fig. 4.21). Similar types of cell death were also found after targeting hsa_circ_0005588 and hsa_circ_0005598 in the DLD-1 cell line although the top essential circRNA, hsa_circ_0005703 affected DLD-1 cells to a lesser extent compared to HCT 116 cells. On the other hand, targeting the intergenic circRNA (hsa_circ_0002989) by the shRNAs did not cause extensive cell death in either cell line.

4.7 Results of Cell Proliferation Assay in Colorectal Cancer Cell Lines

To test whether the knockdown of circRNAs caused any effect on cell proliferation, we used the same colorectal cancer cell lines. We showed that both shRNAs for all four circRNAs (hsa_circ_0005703, hsa_circ_0005588, hsa_circ_0002989, hsa_circ_0005598) affected cell proliferation in both HCT 116 and DLD-1 cell lines in comparison to control shRNA (Fig. 4.22). The growth of HCT 116 cells reached stationary phase on fourth day while DLD-1 cells were still in log phase on the fourth day. Though equal number of cells were seeded for both cell lines on the first day, they grew at different rates probably due to different doubling time and/or varying effect of lentiviral transduction on them. Considering the effect of both shRNAs

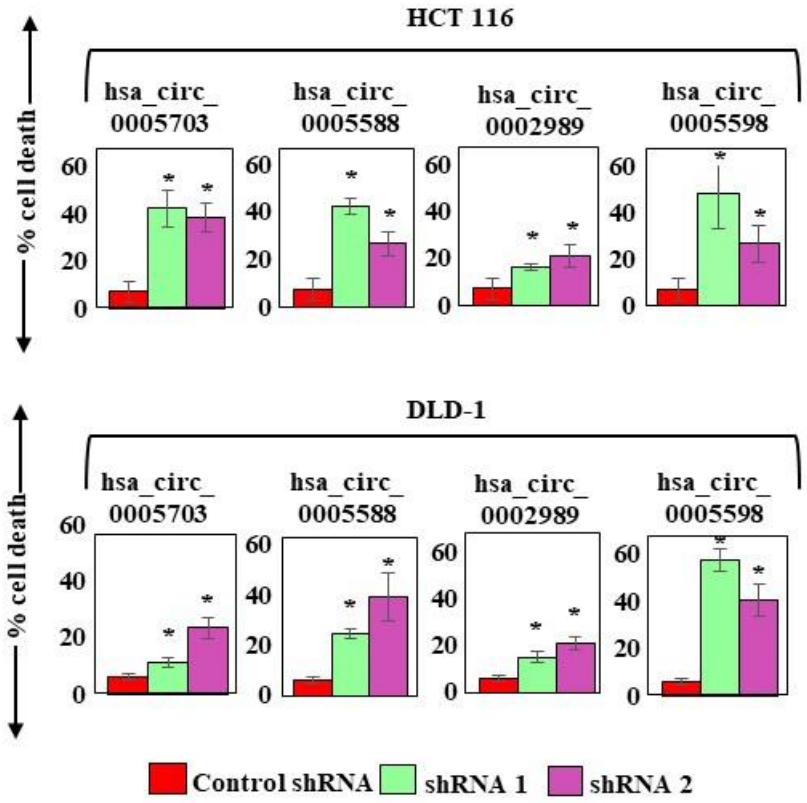

Figure 4.21 Cell death assay for knockdown of top essential circRNAs in colorectal cancer cell lines. Bar diagrams showing comparison of cell death in colorectal cancer cell lines after targeting four different essential circRNAs identified from our screen. Percentages of the cell death are plotted in Y-axis against respective shRNAs plotted in X-axis. Knockdown and matching control cells were seeded and cultured for 72 hours before staining and flow cytometry. The experiments were done in triplicates, means ± standard deviation (error bars). **P* < 0.05 (Student's t-test).

for a circRNA (average of normalized live cell numbers on the fourth day), proliferation of HCT 116 cells showed very similar effect in targeting all of the four circRNAs. But DLD-1 cells showed higher effect in targeting hsa_circ_0005598 in comparison to other three circRNAs. Though targeting hsa_circ_0002989 in both the cell lines did not cause very high cell death, it caused proliferative disadvantages to cells. This indicated that cell proliferation was not only affected by cell death but also slow growth rate. Moreover, though targeting hsa_circ_0005703 in DLD-1 cells caused lesser cell death (than HCT 116) after 24 hours, slower growth was observed after 96 hours. This indicated that cell lines of even the same cancer type responded differently to knockdown of hsa_circ_0005703.

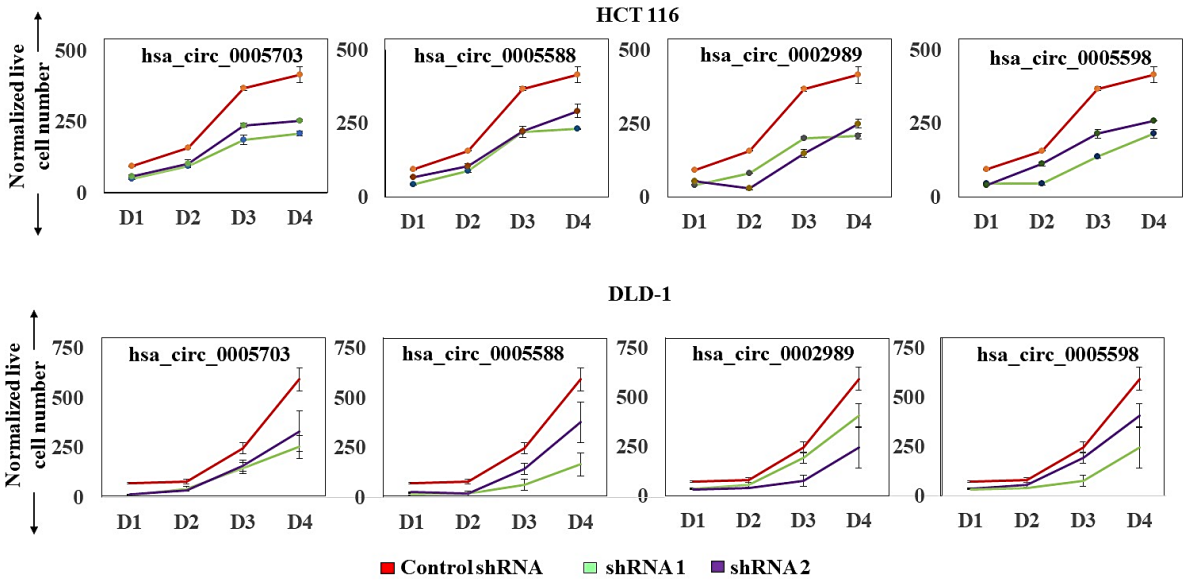

Figure 4.22 Cell proliferation assay for knockdown of top essential circRNAs in colorectal cancer cell lines. Line charts showing proliferation of colorectal cancer cell lines after targeting four different essential circRNAs identified from our screen. Number of days are plotted in X-axis and normalized live cell numbers (for both shRNAs targeting a circRNA) are plotted in Y-axis. Knockdown and matching control cells were seeded and cultured for four days before harvest and counting. The experiments were done in duplicates.

4.8 Results of Colony Formation Assay in Colorectal Cancer Cell Lines

To test whether the shRNA-mediated knockdown of circRNAs caused any effect on colony forming ability of cancer cells, we used two colorectal cell lines (HCT 116 and DLD-1). We observed that both E- and intergenic circRNAs showed significantly less colony formation after targeting them with shRNAs in comparison to control shRNA in both HCT 116 and DLD-1 cell lines (Fig. 4.23). Targeting all four circRNA showed a very similar effect on colony formation and proliferation for HCT 116 cells. DLD-1 cells showed reduced effect in colony formation for targeting hsa_circ_0005588 and hsa_circ_0002989 as compared to the respective effect in proliferation. This indicated that, DLD-1 cells possibly developed resistance mechanism with time to overcome the effect caused by knockdown of respective circRNAs.

4.9 Results of Tumorsphere Assay in Colorectal Cancer Cell Lines

To examine whether the knockdown of circRNAs caused any effect on tumor aggressiveness properties, we used two colorectal cell lines (HCT 116 and DLD-1) for tumorsphere assay. We observed that exonic circRNAs showed significantly less tumorsphere

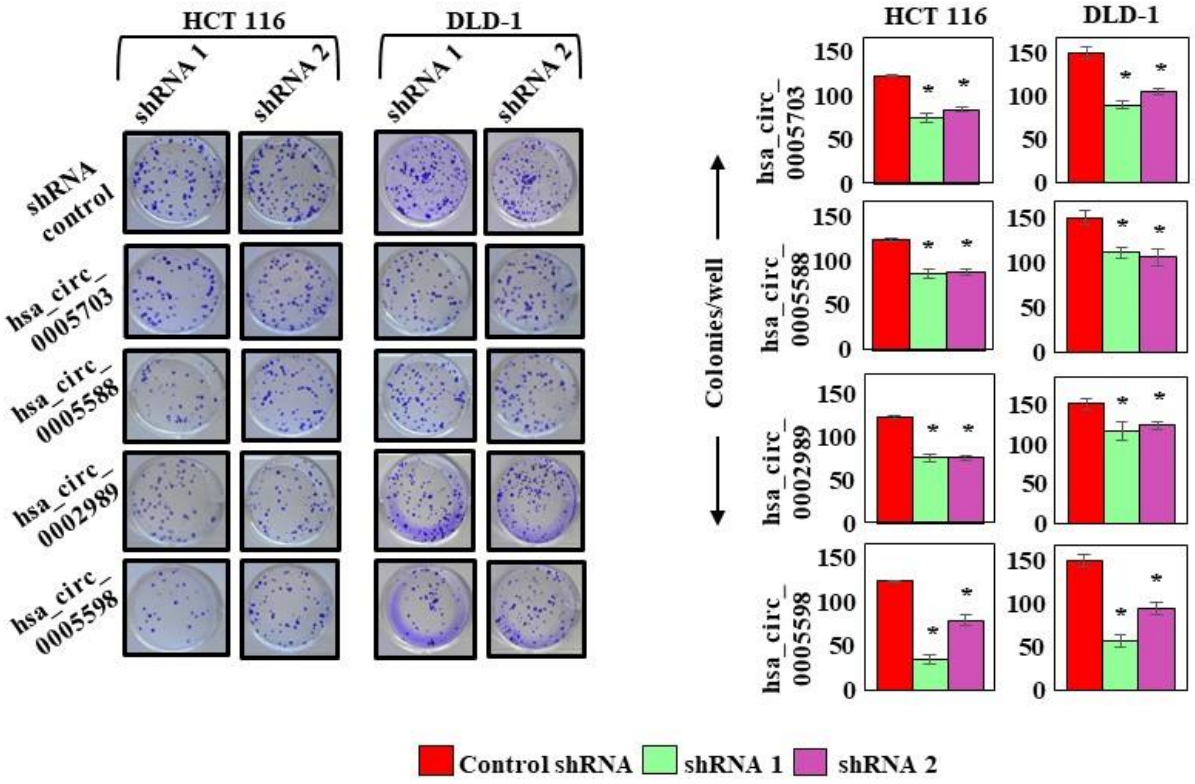

Figure 4.23 Colony formation assay for knockdown of top essential circRNAs in colorectal cancer cell lines. Images and bar diagrams showing comparison of colony formation after targeting four different essential circRNAs identified from our screen. In the bar diagrams, number of colonies are plotted in Y-axis against respective shRNAs plotted in X-axis. Knockdown and matching control cells were seeded and cultured for seven days before colony staining, imaging and counting. The experiments were done in duplicates, means ± standard deviation (error bars). *$P < 0.05$ (Student's t-test).

forming cells after targeting them with shRNAs in comparison to control shRNA in both HCT 116 and DLD-1 cell lines (Fig. 4.24). On the other hand, targeting intergenic circRNA did not affect tumorsphere formation significantly in both cell lines (for two shRNAs in HCT 116 cells and one shRNA in DLD-1 cells).

4.10 Results of Cell Death Assay in Other Cancer Types

We tested the essentiality of the four circRNAs in two breast cancer and two prostate cancer cell lines using 7-AAD staining and flow cytometry. We checked the expression of the parental genes of E-circRNAs from 'CCLE' for all the mentioned cell lines. We found that all

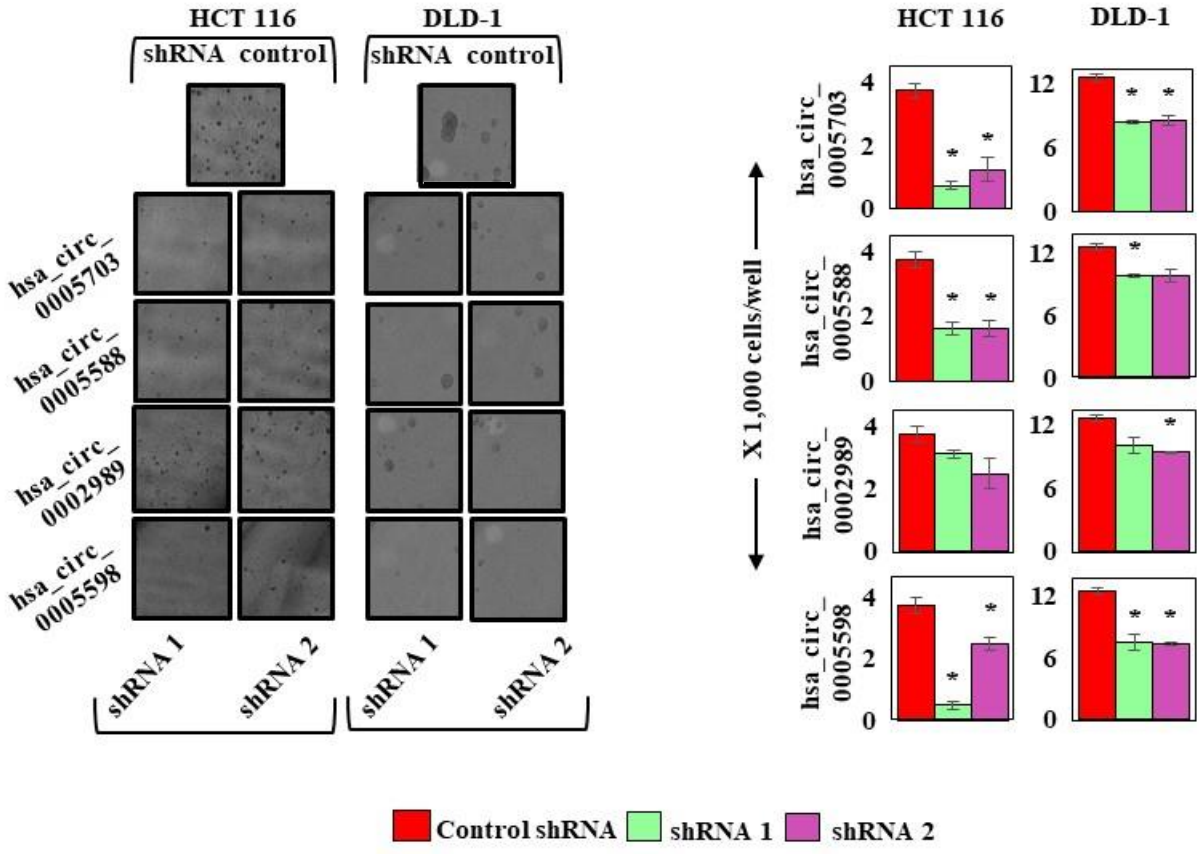

Figure 4.24 Tumorsphere assay for knockdown of circRNAs in colorectal cancer cell lines. Images and bar diagrams showing comparison of tumorsphere forming cells after targeting four different essential circRNAs identified from our screen. In the bar diagrams, number of tumorsphere forming cells are plotted in Y-axis against respective shRNAs plotted in X-axis. For HCT 116 cells, 4 X images are presented while 10 X images are presented for DLD-1 cells. Knockdown and matching control were seeded in ultra-low attachment plates and cultured for seven days. The resulting tumorspheres were collected, dissociated, and the total number of cells were counted. The experiments were done in duplicates, means ± standard deviation (error bars). *$P < 0.05$ (Student's t-test).

of these genes have higher expression level than the median of all human gene expression (Fig 4.25). As both genes and corresponding circRNAs come from the same splicing process, we assumed if the parental gene is expressed in a cell line, there is the possibility of corresponding E-circRNA expression. We found that cell lines from other cancer types did not follow the same essentiality profile like HCT 116 under the same experimental conditions. Breast cancer cell line MDA-MB-231 showed reduced essentiality for all four essential circRNA in comparison to HCT 116 cell line (Fig. 4.26). On the other hand, breast cancer cell line MCF7 exhibited higher essentiality than MDA-MB-231 for all four essential circRNAs. Additionally,

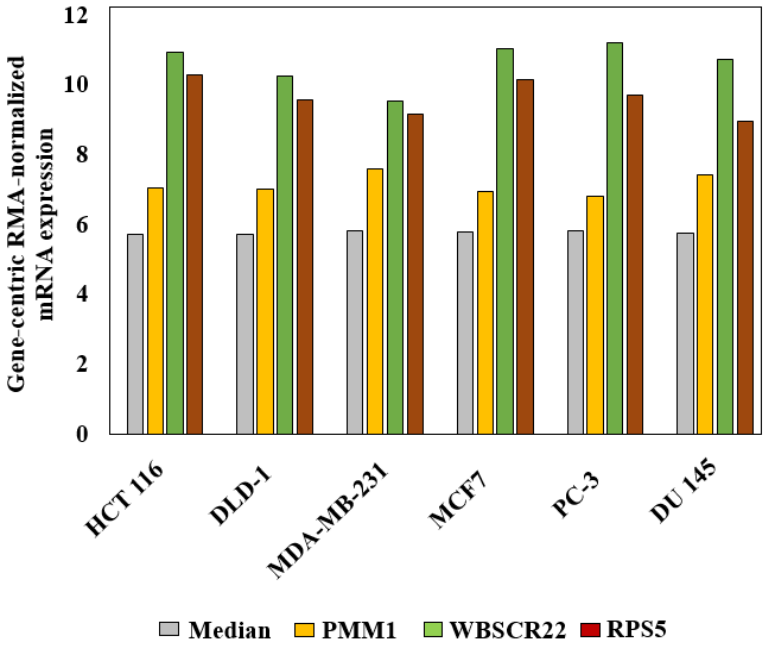

Figure 4.25 Expression study of parental genes of top essential E-circRNAs in different cancer cell lines. (A) Bar diagram showing expression of parental genes of top three essential E-circRNAs from 'CCLE' in HCT 116, DLD-1, MDA-MB-231, MCF7, PC-3 and DU-145 cell lines. The names of the cell lines are plotted in X-axis and expression data are plotted in Y-axis. The expression data provided by 'CCLE' showed mRNA expression levels, which was RMA-normalized gene-centric expression according to 'CCLE' standards.

prostate cancer cell line DU 145 and PC-3 showed reduced essentiality for all of the four circRNAs in comparison to HCT 116 cell line. These results suggested that circRNA essentiality has cell line and/or tissue specificity.

4.11 Predictive Analysis of Circular RNA Interactions with miRNA and Protein

We performed a bioinformatics analysis on 'CircInteractome' to predict miRNA and RNA binding protein (RBP) interaction with the top five essential circRNAs from the screening. We found different circRNAs have binding sites for multiple miRNAs, indicating their possible involvement in diverse gene regulatory processes (Table 4.3). There were almost no common miRNAs among the five circRNAs. Most of the miRNA had only one binding site in the circRNAs. Conversely, hsa_circ_005588 had two binding sites for hsa-miR-338-3p (highlighted with red) and hsa_circ_005588 had two binding sites for hsa-miR-1825 (highlighted with red). In the case of predicted RBP interactions with five circRNAs, hsa_circ_005598 had the highest number of predicted interaction partners and

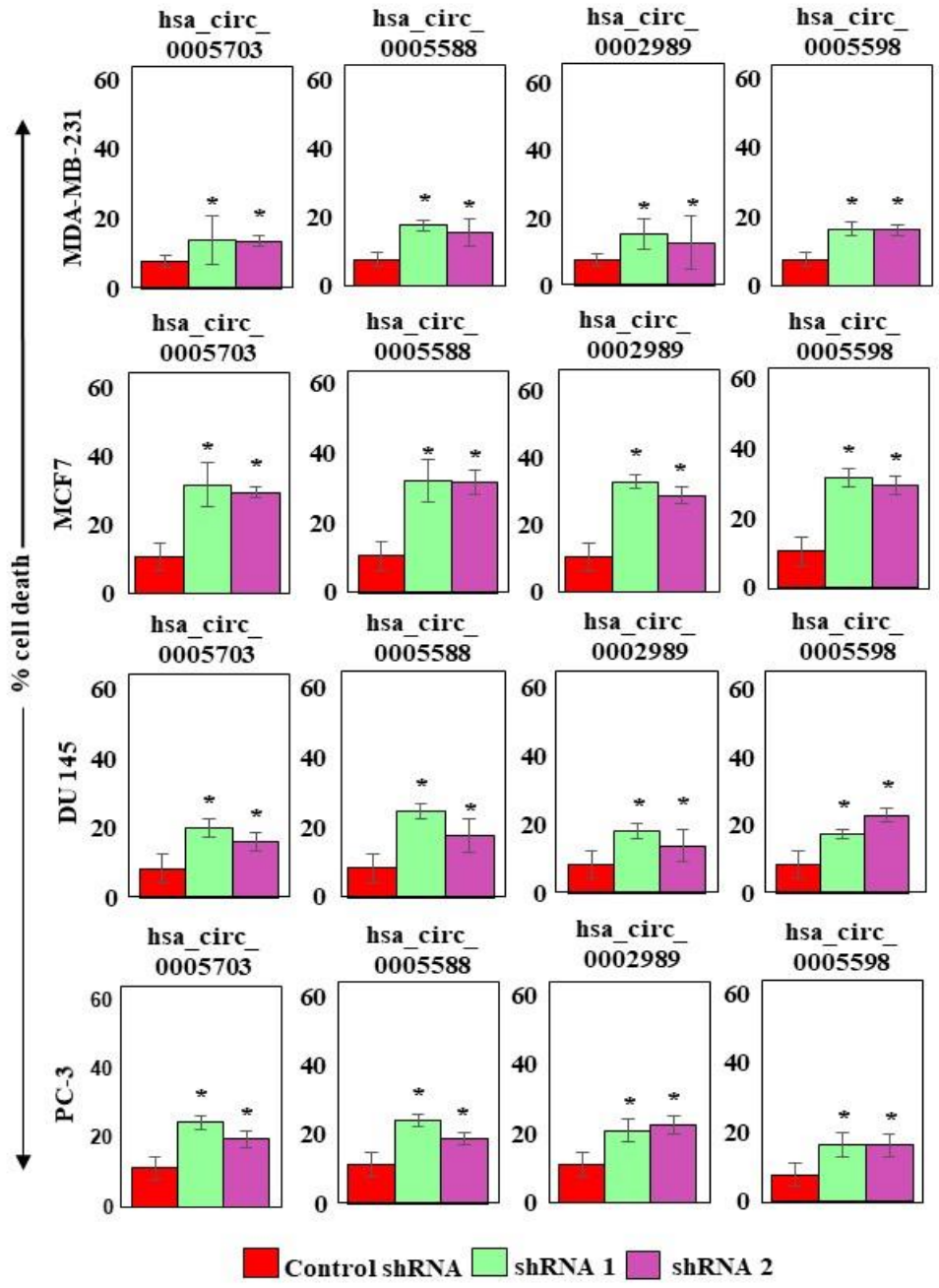

Figure 4.26 Cell line specific essentiality of top essential circRNAs. Bar diagrams showing comparison of cell death in cell lines of breast and prostate cancers after targeting four different essential circRNAs identified from our screen. Percentages of the cell death are plotted in Y-axis against respective shRNAs plotted in X-axis. Knockdown and matching control cells were seeded and cultured for 72 hours before staining and flow cytometry. The experiments were done in triplicates, means ± standard deviation (error bars). *$P < 0.05$ (Student's t-test).

Table 4.3 Predicted interacting miRNAs with top essential circRNAs

hsa_circ_0005703	hsa_circ_0005588	hsa_circ_0092290	hsa_circ_0002989	hsa_circ_0005598
hsa-miR-1184	hsa-miR-1227	hsa-miR-1224-3p	hsa-miR-1231	hsa-miR-1208
hsa-miR-1270	hsa-miR-146b-3p	hsa-miR-1225-3p	hsa-miR-1246	hsa-miR-1289
hsa-miR-1299	hsa-miR-1825	hsa-miR-1228	hsa-miR-1265	hsa-miR-1294
hsa-miR-1307	hsa-miR-186	hsa-miR-1233	hsa-miR-1292	hsa-miR-1827
hsa-miR-145	hsa-miR-194	hsa-miR-1827	hsa-miR-1825	hsa-miR-197
hsa-miR-1827	hsa-miR-326	hsa-miR-187	hsa-miR-197	hsa-miR-377
hsa-miR-198	hsa-miR-330-5p	hsa-miR-512-5p	hsa-miR-224	hsa-miR-490-5p
hsa-miR-217	hsa-miR-338-3p	hsa-miR-636	hsa-miR-338-3p	hsa-miR-492
hsa-miR-361-3p	hsa-miR-361-3p	hsa-miR-644	hsa-miR-346	hsa-miR-517b
hsa-miR-432	hsa-miR-512-5p	hsa-miR-663b	hsa-miR-492	hsa-miR-556-5p

hsa_circ_00092290 had only a single interacting partner named EIF4A3 (Fig 4.27). Some RBPs were common for multiple circRNAs. Interestingly, all of the top five essential circRNAs have multiple binding sites for EIF4A3, indicating a possible involvement of this protein in the essentiality mechanism of these circRNAs.

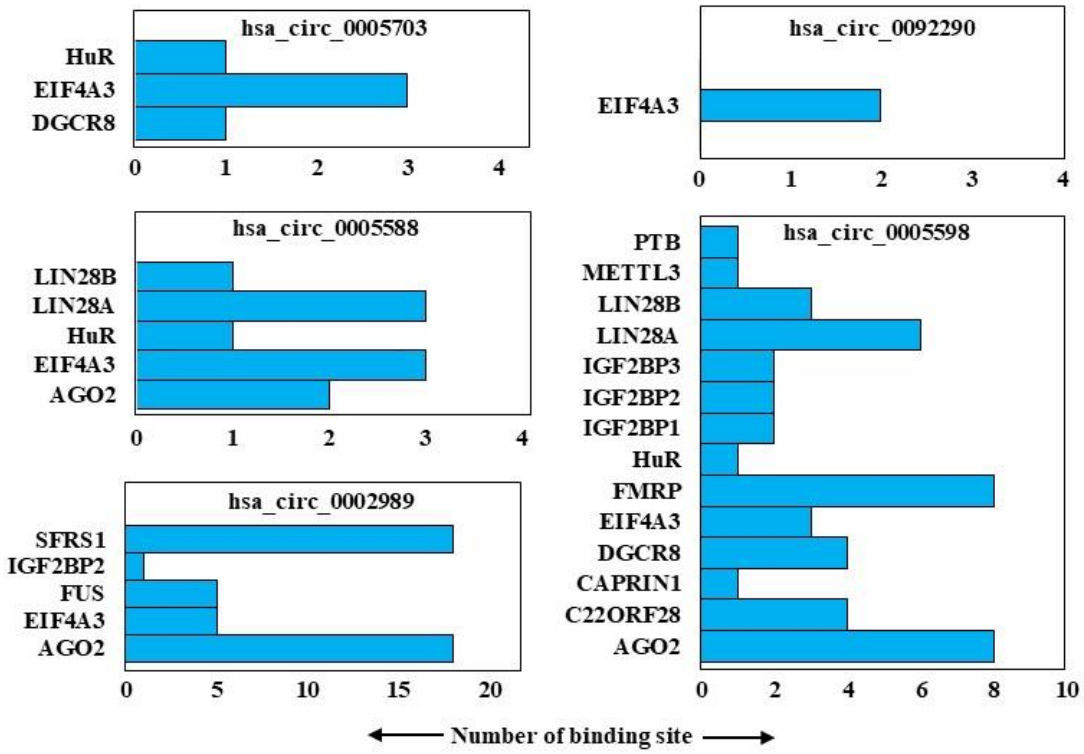

Figure 4.27 Predicting interacting proteins of top essential circRNAs. Bar diagrams showing predicted interacting proteins of top five essential circRNAs. Numbers of predicted binding sites are plotted in X-axis and names of the respective proteins are plotted in Y-axis.

5.0 Discussion

5.1 Development of a Novel Tool and Method for Systematic Study of Circular RNA Essentiality

CircRNA is a new class of RNA molecules discovered by recent advancement in genomic technologies. Many of the circRNAs are found to be involved in gene regulation during development and pathogenesis. Unfortunately, the roles of most circRNAs are not still known, which requires intensive studies to decipher their functional characteristics. Studying functions of each circRNA individually (from the vast repertoire) requires time and resources. From this viewpoint, expression profiling has come into use for studying differential expression of circRNAs in normal and diseased conditions (e.g. cancer) (Li *et al.*, 2015e; Qu *et al.*, 2015b; Ahmed *et al.*, 2016; Song *et al.*, 2016). Though expression studies can help in biomarker discovery, there are some major pitfalls, as expression is not always a cause of a biological or molecular phenomenon. Specifically, from therapeutic point of view, it is important to identify the causative factor necessary for particular disease phenotypes. In this regard, genome-scale, pooled shRNA screening is a suitable way to investigate large set of genes and/or transcripts for a particular characteristic (e.g. gene essentiality). From this perspective, we decided to screen a significant number of circRNAs identified from different genomic studies for their essentiality in cancer cells. Currently, there are no commercial tools (e.g. shRNA library, CRISPR library) available for such studies. As circRNAs are alternative products generated during the splicing process, they cannot be targeted by more efficient screening tools like CRISPR library, as it targets DNA element rather than RNA transcript. To address these challenges, we developed a novel ~ 15K shRNA library, which specifically targets ~ 5K circRNAs. We effectively established a complete pipeline for design, construction, and validation of a circRNA shRNA library. We used the circRNA shRNA library to systematically screen a colorectal cancer cell line (HCT 116) and identify essential circRNAs using a lentivirus-based pooled screening method. For successful identification of essential circRNAs, we developed a scoring algorithm to analyze NGS results from the screening procedure. Such large-scale, genome-wide screens have been applied to identify cancer-specific essential genes in some studies (Paddison *et al.*, 2004; Luo *et al.*, 2008; Schlabach *et al.*, 2008; Silva *et al.*, 2008; Barbie *et al.*, 2009; Marcotte *et al.*, 2012; Koike-Yusa *et al.*, 2013; Shalem *et al.*, 2014; Wang *et al.*, 2014b; Hart *et al.*, 2015; Munoz *et al.*, 2016). Moreover, some research also utilized genome-wide screening for

detecting therapeutically relevant synthetic lethal targets (Luo *et al.*, 2009; Brough *et al.*, 2011; Vizeacoumar *et al.*, 2013; Bajrami *et al.* 2014; Cermelli *et al.*, 2014; Paul *et al.*, 2014; Van Der Meer, *et al.*, 2014; Paul *et al.*, 2016).

5.2 Improvement of shRNA and CRISPR Library Sequencing Methods on Next-Generation Sequencing Platform

Direct shRNA library sequencing from genomic DNA generally results in polyclonal, low quality, and incomplete reads due to hairpin and/or heteroduplex formation. This challenge is significantly amplified in low-to-medium throughput bench-top sequencers as they already have limited coverage of sequencing. Ideally, shRNA libraries are designed to form functional hairpin structures when processed within the cell. Because of this, it is natural that these sequences form hairpin structures during sequencing reactions and reduce sequencing quality, irrespective of the sequencing platform that is being used. Similarly, heteroduplex structures can also reduce sequencing quality. This heteroduplex issue is not unique to shRNA libraries alone but affects all types of mixed-oligo libraries (e.g. CRISPR, phage display libraries, etc.). For example, one study showed that 70% of sequences could not be read due to heteroduplex formation in a phage-selected peptide library (Rebollo *et al.*, 2014). Similarly, 40% of reads from Illumina sequencing of genetically distinct HIV-1 genome were unexpected recombinant sequences due to heteroduplex formation (Liu *et al.*, 2014). This heteroduplex issue also makes it very challenging when sequencing larger libraries, as it hinders multiplexing (running more samples at a time for sequencing). Unfortunately, computational predictions to simulate the formation of heteroduplex structures are also limited due to high randomness associated with the formation of these structures.

Our method aims to eliminate sequence failure and maximize throughput in low-to-medium throughput bench-top sequencers. We showed that the adapter-ligated half-shRNA sequencing increased usable read output by ~ 60%, relative to current sequencing strategies. To confirm that our experimental procedures successfully eliminated the hairpin related issues, we carried out sequencing of both full-shRNA and half-shRNA library samples. We showed that sequencing half-shRNA reduced the polyclonal, low quality, and sequence termination while increasing reads of the intended library. In addition, reduction of heteroduplex structures also maximized our throughput. As increased-fold representation of the shRNA library is important

for the reproducibility of pooled shRNA screens (Strezoska *et al.*, 2012), our methodology should alleviate any of these concerns because sequencing of half-hairpins increased the fold coverage by at least five times.

Our method resulted in improvements of the initial quality assessment by the server software (after loading and enrichment) in Ion Torrent sequencing. While shRNA sequencing showed 21% initial polyclonal reads and 45% low quality reads, half-shRNA sequencing reduced them to 12% and 36%, respectively. After heteroduplex removal, half-shRNA sequencing decreased the polyclonal and low quality reads further to 6.5% and 9%, respectively. In comparison to some of the established methods such as MuPlus (transposon-based), MuSeek (commercial), and other ligation-based methods, our approach showed a significant reduction in the rate of polyclonal reads. While these methods exhibited 23%, 51%, and 31% polyclonal reads, respectively, our method produced only 6.5% (Gorbacheva, *et al.*, 2015). Our method also reduced low quality reads (9%) in comparison to the commercial MuSeek method (55%), although MuPlus and other ligation-based methods have similar levels of efficiency (MuPlus: 6% and other ligation-based: 9%) (Gorbacheva *et al.*, 2015). Additionally, final sequence output after loading, enrichment, initial polyclonal removal, and low-quality removal have been significantly increased with our method (88%), when compared to other methods (MuPlus: 72%, MuSeek: 22% and other ligation-based method: 63%) (Gorbacheva *et al.*, 2015). Unlike previously used sequencing strategies (Hoshiyama *et al.*, 2012), our ligation-based, half-shRNA sequencing is readily amenable to any pooled shRNA screening studies. Our strategy to detect and minimize heteroduplex formation in shRNA or CRISPR libraries can also be extended to any mixed-oligo libraries.

As targeted screening by small coverage libraries, using low-to-medium throughput bench-top sequencers demand minimal unusable reads that arise from hairpin and/or heteroduplex formation, we expect that our approach will be beneficial. Our method can also be extended to whole genome screening using Illumina platforms. With the recent increase in the library sizes such as the ultracomplex-pooled shRNA libraries with 25 shRNAs per gene (Bassik *et al.*, 2013) or the TKO library, where 10 guide RNAs target each gene (Hart, 2015), there is an ever-increasing demand on the sequencing throughput. Therefore, in its current state, maximizing the sequencing strategy is becoming a pressing issue. A whole genome screen using a 90K shRNA pool may require ~ 45 million reads for ~ 500-fold coverage for a

single time point. In fact, a single screen requires multiple time points and replicates, easily exceeding the limitations of a bench-top sequencer (e.g. Ion PGM™ System). Though Illumina technologies provides vast coverage of sequencing, over 40% to 50% of the reads are still non-usable. We expect that our methodological improvements will minimize the loss of reads and maximize throughput and multiplexing capabilities.

5.3 Identification and Characterization of Unique Essential Circular RNA Molecules in Cancer Cells

Our screening of the HCT 116 cell line and subsequent analysis of NGS data identified ~100 essential circRNA hits which might be very critical for the survival of this colorectal cancer cell. As genome-wide screening procedure generates some false positive hits, we selected the top five essential circRNAs from our analysis for further validation. We successfully cloned two shRNA oligonucleotides targeting the top four essential circRNAs (three E-circRNA and one intergenic circRNA) into the pLKO.1 plasmid. We used lentiviral method to target these four circRNAs in HCT 116 cell line. We confirmed knockdown of these circRNAs as well by designing specific divergent primers. For three E-circRNAs, we also confirmed that our shRNAs only targeted the circRNAs and not the corresponding mRNAs. We observed different cellular effects due to the knockdown of the mentioned circRNAs with cell death assay, proliferation assay, colony formation assay and tumorsphere assay in HCT 116 and DLD-1 cell lines. Our results of the cell death assay showed that targeting the top essential circRNAs caused significant amount of cell death in two colorectal cancer cell lines. Proliferation assay and colony formation assay confirmed that knockdown of the essential circRNAs also disrupted normal proliferative properties of the cells. Furthermore, tumorsphere assay showed that knockdown of some of the essential circRNAs altered tumor aggressive properties of the cells. With cell death assay in two breast cancer and two prostate cancer cell lines**,** we also observed that essentiality of the mentioned circRNAs had cell line specific nature.

These essential circRNAs identified from our study have not been previously reported. Moreover, their expression profiles in cancer have not been described by any other authors. There is evidence that some circRNAs can be upregulated or downregulated in different types of cancer (Li. *et al.*, 2015c; Wang *et al.* 2015; Qin *et al.*, 2016; Sand *et al.*, 2016; Xuan *et al.*,

2016; Weng *et al.*, 2017). To have a better understanding, we studied parental genes of the top ~100 E-circRNAs from our list and found that some of them are upregulated in colorectal cancer in comparison to the normal tissues while some are downregulated. Interestingly, these genes had different types of essentiality that was identified from a previous screening (Hart *et al.*, 2015). For example, *PMM1* (parental gene of hsa_circ_005703) is downregulated in cancer and has very low essentiality score (indicating very less involvement in cancer cell survival). Both *WBSCR22* (parental gene of hsa_circ_005588) and *RPS5* (parental gene of hsa_circ_005598) are upregulated in colorectal cancer and have a very high essentiality score (indicating higher importance in cancer cell survival). We assumed that the E-circRNAs can be essential irrespective of the expression and essentiality of the parental genes.

5.4 Building a Foundation for Further Study on Mechanistic Understanding of Circular RNA Essentiality

Our study set a base for mechanistic understanding of circRNA essentiality. Two models can be postulated for the circRNA essentiality. The first model can be described by circRNA and RBP interaction. For example, hsa_circ_005703 had three predicted binding sites for EILF4A3 (eukaryotic translation initiation factor 4A3). Essential circRNAs can bind to three different EILF4A3 proteins and this assembly can help to initiate the translation process. Knockdown of circRNAs can disrupt this interaction, as well as affect cell cycle progression resulting in decreased cancer cell proliferation or increased cell death (Fig 5.1A). Some similar types of mechanisms were also described by other groups (Du *et al.*, 2016; Yang, 2016; Du *et al.*, 2017). A second model can be described by sponging of miRNAs by circRNA. Essential circRNAs can sponge AGO-miRNA complexes. This active sponging of miRNAs, can cause excessive mRNA levels of some oncogenes resulting in enhanced proliferation of cancer cells. Knockdown of essential circRNAs can release AGO-miRNA complexes which can repress the oncogenes, resulting in decreased cancer cell proliferation or increased cell death (Fig. 5.1B). For example, hsa_circ_005588 has two target sites for miR-338-3p which was found to be a highly conserved miRNA identified from 'TargetScan' analysis. Furthermore, this miRNA is an established tumor suppressor (Xue *et al.* 2014) and has some oncogenic targets (predicted from 'ToppGene' analysis). For instance, *CDK4* (Cyclin-dependent kinase 4), an established oncogene (https://www.ncbi.nlm.nih.gov/gene: NCBI) and cell cycle regulatory protein, is a

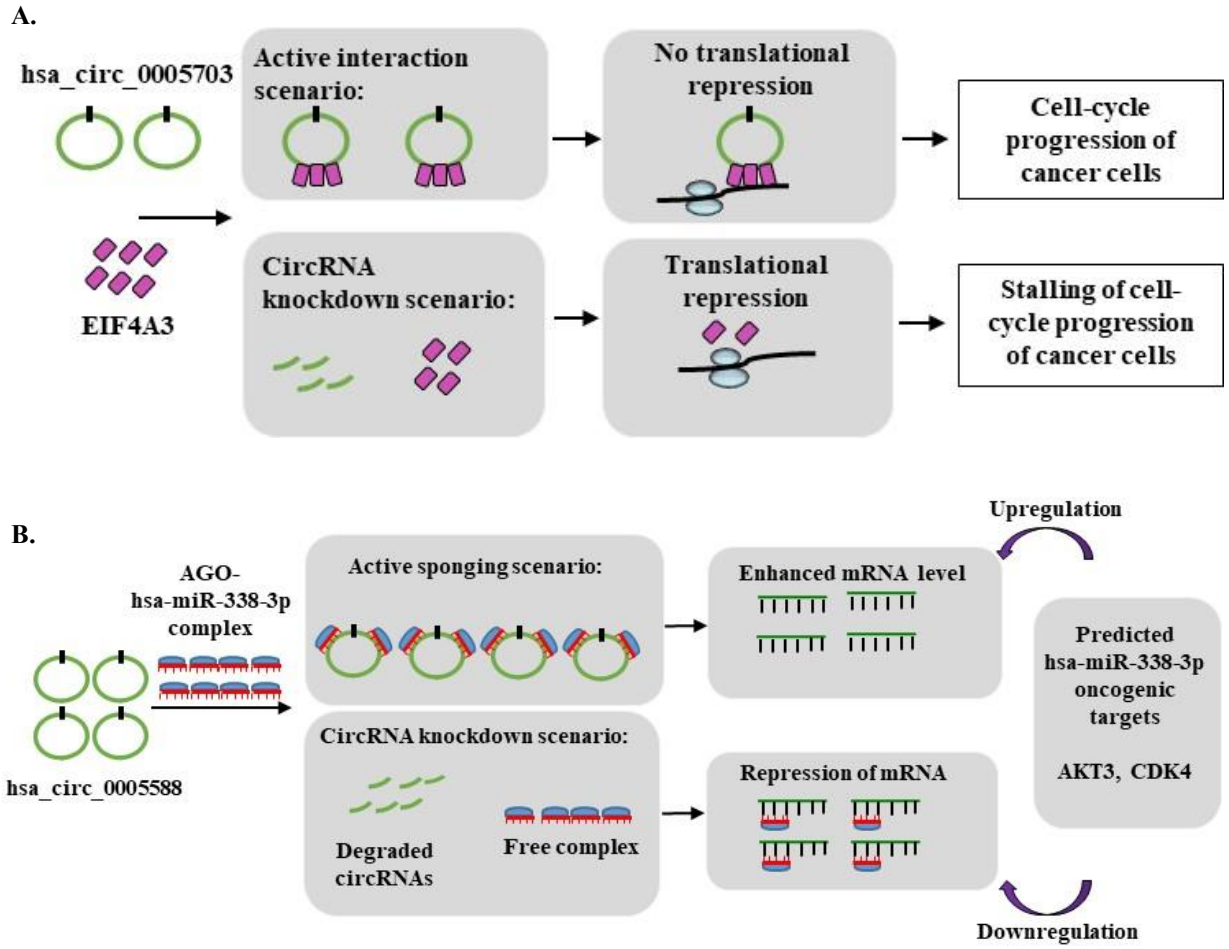

Figure 5.1 Model of circRNA essentiality mechanism. **(A)** Illustration of interaction between circRNAs and EIF4A3 and its role in cell-cycle progression. **(B)** Illustration describing the model of has-miR-338-3p sponging by has_circ_0005588 and its role in cell-cycle progression.

target of miR-338-3p. Though circRNAs having only one or two miRNA sites cannot be supported strongly by the sponging model, the relative abundance of circRNAs due to their stability might ensure their higher copy numbers in cancer cells and thus impart functional effect in cellular phenotypes (Salzman *et al.*, 2012; Salzman *et al.*, 2013; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Rybak-Wolf *et al.*, 2015). Apart from circRNA based gene regulation mediated by miRNA sponging, E-circRNAs might have some roles in regulating their parental genes to enhance cancer progression. For instance, cancer cells might maintain a ratio of circRNA and mRNA (for exonic circRNAs) to create a competition for tumor suppressor miRNAs between circRNA and mRNA. This is particularly reasonable when the parental gene

is oncogenic and upregulated in cancer. A similar type of mechanism was described by Zhong *et al.*, 2016. Analysis of 'CCSD' database also showed that a ratio of circRNA and mRNA originally exists in different cancer samples. For example, expression ratio of hsa_circ_005588/WBSCR22 and hsa_circ_005598/RPS5 were found to be 0.58 and 0.50 respectively in K-562 cell line (leukemia). On the other hand, circRNAs more likely regulate any other genes except the parental genes, specifically when the parental genes are downregulated.

Based on the predicted interactions of circRNA with proteins, miRNAs and long non-coding lincRNAs, cancer specific interactome can be built for circRNAs. An example of such breast cancer interactome for CDR1as was generated from 'circ2traits' database as shown in Fig. 5.2. Such interactome study can also be performed for essential circRNAs identified from genomic screens. Thus it is possible to bridge the essential circRNAs to global cancer interactome.

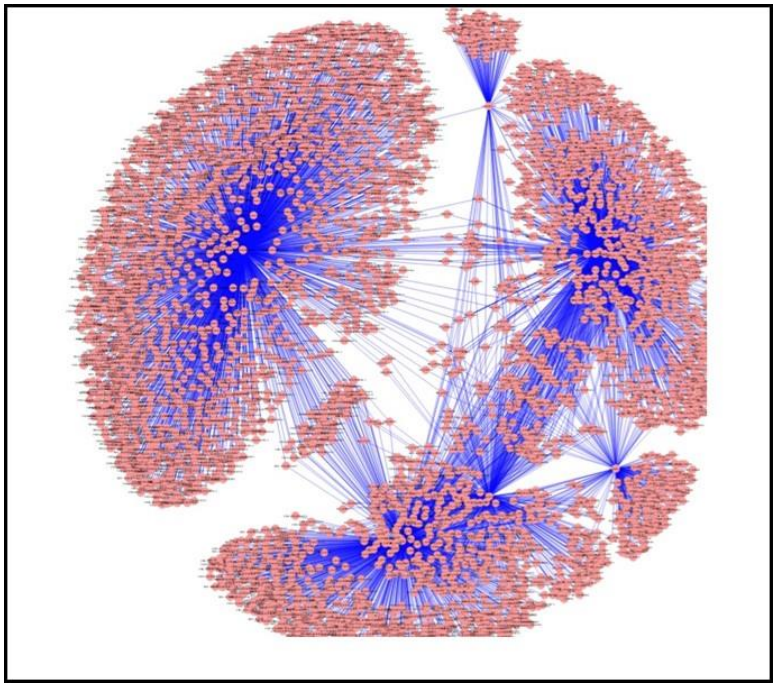

Figure 5.2 CircRNA interactome study. Predicted interactome of CDR1as in breast cancer indicating a model of circRNA specific gene regulatory network. The regulatory network that comprises of circRNAs, miRNAs, mRNAs and lncRNAs, is also applicable in integrating essential circRNAs into the existing cancer knowledgebase.

6.0 Conclusions and Future Directions

6.1 Conclusions

CircRNAs are a newly discovered interesting class of RNA molecules. There are limitations of the existing knowledge about their roles in complex human disease like cancer. In this study, we took a novel approach to investigate the essentiality of many of these circRNAs with a perspective to understand their roles in cancer biology. We developed a circRNA specific shRNA library to systematically screen these circRNAs for their essentiality in cancer. After targeting colorectal cancer cell line with the library through a standard screening procedure, several essential circRNA hits were identified using a novel scoring method. The top four essential circRNAs were then validated further with some expression and cell based assays. Moreover, expression and enrichment studies were performed on the parental genes of corresponding E-circRNAs to understand their functional relevance to cancer. Finally, we also performed some bioinformatics study to propose two models of circRNA-associated essentiality mechanism.

6.2 Future Directions

To know the involvement of the essential circRNAs with different biological pathways, high-throughput pathway enrichment studies can be performed. Epigenetic changes of circRNAs and their relevance to essentiality can lead to a novel field such as circRNA epigenomics. Moreover, genome wide association study of essential circRNAs can be effective to understand the disease phenotypes. As all the works were performed *in vitro* model so far, the targets should also be tested in appropriate mouse model to enhance the reliability of the circRNA essentiality. Given that the expression profiling has been completed for circRNAs between normal and cancer tissues, they can be also used to identify synthetic lethal or synthetic dosage lethal partners for therapeutic intervention. Specifically, highly downregulated circRNAs in cancer can be used to find synthetic lethal partners while highly upregulated circRNAs in cancer can be used to find synthetic dosage lethal partners through similar type of genome-wide screening. As current studies suggest that circRNAs can be translated, it may be interesting to investigate whether essential circRNAs identified from our study can be translated into proteins. The vast repertoire of stable circRNAs can be tested for novel biomarker discovery, particularly those with differential expression. Furthermore, cancer

specific circRNA essentiality can be applied for therapeutic intervention of cancer. Antisense therapy may be a better approach in this regard though its success is dependent on efficient drug delivery system into the cells. As a whole, circRNA field has remarkable prospects in understanding cancer in a more mechanistic way and utilizing the knowledge for probable therapeutic and diagnostic advancement. Our study provides a foundation for new studies to understand the roles of circRNAs in cancer. Most importantly, our established circRNA shRNA library is a useful tool that can be used further to screen essential circRNAs in other *in vitro* or *in vivo* models. We screened a single cancer cell line for circRNA essentiality, representing a particular type of cancer. But there is ample opportunity to broaden the scope of such study by screening diverse cancer cell lines that cover all major types of malignancies. This is particularly significant for developing a global circRNA essentiality map. The essentiality study of circRNAs can be coupled with their differential expression profiling between normal and cancer tissues. This will help to identify more effective circRNAs from both a therapeutic and diagnostic point of view. Clinical patient derived data from public repository (e.g. 'TCGA') can be very useful in this regard. We validated the essentiality of only four circRNAs from our essentiality list and many others are still remaining for further validation. Experimental validation of the essentiality mechanisms of circRNAs is another vital direction of the circRNA field. Particularly, more intronic and intergenic circRNAs should be considered with importance for future validation and mechanistic study. We discarded all the circRNAs having single shRNA in the library during data analysis although these may be interesting targets as well for future studies. Mechanistic study of E-circRNAs might also help understand many unknown functions of their parental genes.

7.0 References

Ahmed, I., Karedath, T., Andrews, S.S., Al, I.K., Mohamoud, Y.A., Querleu, D., Rafii, A., and Malek, J.A. (2016). Altered expression pattern of circular RNAs in primary and metastatic sites of epithelial ovarian carcinoma. Oncotarget *7*, 36366-36381.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. Mol. Cell *56*, 55-66.

Bachmayr-Heyda, A., Reiner, A.T., Auer, K., Sukhbaatar, N., Aust, S., Bachleitner-Hofmann, T., Mesteri, I., Grunt, T.W., Zeillinger, R., and Pils, D. (2015). Correlation of circular RNA abundance with proliferation–exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. Sci. Rep. *5*, 8057.

Bajrami, I., Frankum, J.R., Konde, A., Miller, R.E., Rehman, F.L., Brough, R., Campbell, J., Sims, D., Rafiq, R., and Hooper, S. (2014). Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity. Cancer Res. *74*, 287-297.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., and Scholl, C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108-112.

Barrett, S.P., and Salzman, J. (2016). Circular RNAs: analysis, expression and potential functions. Development *143*, 1838-1847.

Barrett, S.P., Wang, P.L., and Salzman, J. (2015). Circular RNA biogenesis can proceed through an exon-containing lariat precursor. Elife *4*, e07540.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., and Mann, M. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell *152*, 909-922.

Blakely, K., Ketela, T., and Moffat, J. (2011). Pooled lentiviral shRNA screening for functional genomics in mammalian cells. Methods Mol. Biol. *781*, 161-182.

Bowman, S.K., Simon, M.D., Deaton, A.M., Tolstorukov, M., Borowsky, M.L., and Kingston, R.E. (2013). Multiplexed Illumina sequencing libraries from picogram quantities of DNA. BMC Genomics *14*, 466.

Brandariz-Fontes, C., Camacho-Sanchez, M., Vila, C., Vega-Pla, J.L., Rico, C., and Leonard, J.A. (2015). Effect of the enzyme and PCR conditions on the quality of high-throughput DNA sequencing results. Sci. Rep. *5*, 8056.

Brough, R., Frankum, J.R., Costa-Cabral, S., Lord, C.J., and Ashworth, A. (2011). Searching for synthetic lethality in cancer. Curr. Opin. Genet. Dev. *21*, 34-41.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell *73*, 1019-1030.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., and Steinmetz, L.M. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393-1406.

Cermelli, S., Jang, I.S., Bernard, B., and Grandori, C. (2014). Synthetic lethal screens as a means to understand and treat MYC-driven cancers. Cold Spring Harb. Perspect. Med. *4*, a014209.

Chen, C.-y., and Sarnow, P. (1995). Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. Science *268*, 415-417.

Chen, X., Han, P., Zhou, T., Guo, X., Song, X., and Li, Y. (2016a). circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. Sci. Rep. *6*, 34985.

Chen, Y., Li, C., Tan, C., and Liu, X. (2016b). Circular RNAs: a new frontier in the study of human diseases. J Med. Genet. *53*, 359-365.

Conn, S.J., Pillman, K.A., Toubia, J., Conn, V.M., Salmanidis, M., Phillips, C.A., Roslan, S., Schreiber, A.W., Gregory, P.A., and Goodall, G.J. (2015). The RNA binding protein quaking regulates formation of circRNAs. Cell *160*, 1125-1134.

Du, W.W., Fang, L., Yang, W., Wu, N., Awan, F.M., Yang, Z., and Yang, B.B. (2017). Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. Cell Death Differ. *24*, 357-370.

Du, W.W., Yang, W., Liu, E., Yang, Z., Dhaliwal, P., and Yang, B.B. (2016). Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. Nucleic Acids Res. *44*, 2846-2858.

Dudekula, D.B., Panda, A.C., Grammatikakis, I., De, S., Abdelmohsen, K., and Gorospe, M. (2016). CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. RNA Biol. *13*, 34-42.

Ghosal, S., Das, S., Sen, R., Basak, P., and Chakrabarti, J. (2013). Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. Front. Genet. *4*, 283.

Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. RNA *20*, 1666-1670.

Gorbacheva, T., Quispe-Tintaya, W., Popov, V.N., Vijg, J., and Maslov, A.Y. (2015). Improved transposon-based library preparation for the Ion Torrent platform. Biotechniques *58*, 200-202.

Guarnerio, J., Bezzi, M., Jeong, J.C., Paffenholz, S.V., Berry, K., Naldini, M.M., Lo-Coco, F., Tay, Y., Beck, A.H., and Pandolfi, P.P. (2016). Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. Cell *165*, 289-302.

Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. Genome Biol. *15*, 409.

Han, Y.-N., Xia, S.-Q., Zhang, Y.-Y., Zheng, J.-H., and Li, W. (2017). Circular RNAs: A novel type of biomarker and genetic tools in cancer. Oncotarget *8*, 64551-64563.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384-388.

Hansen, T.B., Venø, M.T., Damgaard, C.K., and Kjems, J. (2015). Comparison of circular RNA prediction tools. Nucleic Acids Res. *44*, e58-e58.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., and Sun, S. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell *163*, 1515-1526.

Hentze, M.W., and Preiss, T. (2013). Circular RNAs: splicing's enigma variations. The EMBO J. *32*, 923-925.

Hoshiyama, H., Tang, J., Batten, K., Xiao, G., Rouillard, J.-M., Shay, J.W., Xie, Y., and Wright, W.E. (2012). Development of methods for quantitative comparison of pooled shRNAs by mass sequencing. J. Biomol. Screen *17*, 258-265.

Hsu, M.-T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. Nature *280*, 339-340.

Huang, G., Zhu, H., Shi, Y., Wu, W., Cai, H., and Chen, X. (2015). cir-ITCH plays an inhibitory role in colorectal cancer by regulating the Wnt/β-catenin pathway. PloS One *10*, e0131225.

Islam, M.F., Watanabe, A., Wong, L., Lazarou, C., Vizeacoumar, F.S., Abuhussein, O., Hill, W., Uppalapati, M., Geyer, C.R., and Vizeacoumar, F.J. (2017). Enhancing the throughput and multiplexing capabilities of next generation sequencing for efficient implementation of pooled shRNA and CRISPR screens. Sci. Rep. *7*, 1040.

Ivanov, A., Memczak, S., Wyler, E., Torti, F., Porath, H.T., Orejuela, M.R., Piechotta, M., Levanon, E.Y., Landthaler, M., and Dieterich, C. (2015). Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep. *10*, 170-177.

Jalali-Yazdi, F., Huong Lai, L., Takahashi, T.T., and Roberts, R.W. (2016). High-Throughput Measurement of Binding Kinetics by mRNA Display and Next-Generation Sequencing. Angew. Chem. Int. Ed. Engl. *55*, 4007-4010.

Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. RNA *19*, 141-157.

Ketela, T., Heisler, L.E., Brown, K.R., Ammar, R., Kasimer, D., Surendra, A., Ericson, E., Blakely, K., Karamboulas, D., and Smith, A.M. (2011). A comprehensive platform for highly multiplexed mammalian functional genetic screens. BMC Genomics *12*, 213.

Kieleczawa, J. (2005). Simple modifications of the standard DNA sequencing protocol allow for sequencing through siRNA hairpins and other repeats. J. Biomol. Tech. *16*, 220-223.

Kieleczawa, J. (2006). Fundamentals of sequencing of difficult templates—an overview. J. Biomol. Tech. *17*, 207-217.

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M.D.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267-273.

Kristensen, L.S., Okholm, T.L.H., Venø, M.T., and Kjems, J. (2018). Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. RNA Biol. *15*, 280-291.

Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. RNA *20*, 1829-1842.

Lee, Y.S., and Dutta, A. (2009). MicroRNAs in cancer. Annu. Rev. Pathol. *4*, 199-227.

Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., and Dai, L. (2015a). Exon-intron circular RNAs regulate transcription in the nucleus. Nat. Struct. Mol. Biol. *22*, 256-264.

Li, F., Zhang, L., Li, W., Deng, J., Zheng, J., An, M., Lu, J., and Zhou, Y. (2015b). Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/β-catenin pathway. Oncotarget *6*, 6001-60013.

Li, J., Yang, J., Zhou, P., Le, Y., Zhou, C., Wang, S., Xu, D., Lin, H.-K., and Gong, Z. (2015c). Circular RNAs in cancer: novel insights into origins, properties, functions and implications. Am. J. Cancer Res. *5*, 472-480.

Li, P., Chen, S., Chen, H., Mo, X., Li, T., Shao, Y., Xiao, B., and Guo, J. (2015d). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. Clin. Chim Acta *444*, 132-136.

Li, Y., Zheng, Q., Bao, C., Li, S., Guo, W., Zhao, J., Chen, D., Gu, J., He, X., and Huang, S. (2015e). Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. Cell Res. *25*, 981-984.

Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. Genes Dev. *28*, 2233-2247.

Liu, Y.-C., Li, J.-R., Sun, C.-H., Andrews, E., Chao, R.-F., Lin, F.-M., Weng, S.-L., Hsu, S.-D., Huang, C.-C., and Cheng, C. (2015). CircNet: a database of circular RNAs derived from transcriptome sequencing data. Nucleic Acids Res. *44*, D209-D215.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Matochko, W.L., and Derda, R. (2015). Next-generation sequencing of phage-displayed peptide libraries. Methods Mol. Biol. *1248*, 249-266.

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

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., and Munschauer, M. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333-338.

Memczak, S., Papavasileiou, P., Peters, O., and Rajewsky, N. (2015). Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. PloS One *10*, e0141214.

Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb. prot5448.

Miyagishi, M., Sumimoto, H., Miyoshi, H., Kawakami, Y., and Taira, K. (2004). Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. J. Gene Med. *6*, 715-723.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., and Grenier, J.K. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.

Munoz, D.M., Cassiani, P.J., Li, L., Billy, E., Korn, J.M., Jones, M.D., Golji, J., Ruddy, D.A., Yu, K., and McAllister, G. (2016). CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. Cancer Discov. *6*, 900-913.

Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W., and Vogelstein, B. (1991). Scrambled exons. Cell *64*, 607-613.

Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'shaughnessy, A., Gnoj, L., and Scobie, K. (2004). A resource for large-scale RNA-interference-based screens in mammals. Nature *428*, 427-431.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., and Ramberger, E. (2017). Translation of circRNAs. Mol. Cell *66*, 9-21. e27.

Paul, J.M., Templeton, S.D., Baharani, A., Freywald, A., and Vizeacoumar, F.J. (2014). Building high-resolution synthetic lethal networks: a 'Google map'of the cancer cell. Trends Mol. Med. *20*, 704-715.

Paul, J.M., Toosi, B., Vizeacoumar, F.S., Bhanumathy, K.K., Li, Y., Gerger, C., El Zawily, A., Freywald, T., Anderson, D.H., and Mousseau, D. (2016). Targeting synthetic lethality between the SRC kinase and the EPHB6 receptor may benefit cancer treatment. Oncotarget *7*, 50027–50042.

Qin, M., Liu, G., Huo, X., Tao, X., Sun, X., Ge, Z., Yang, J., Fan, J., Liu, L., and Qin, W. (2016). Hsa_circ_0001649: a circular RNA and potential novel biomarker for hepatocellular carcinoma. Cancer Biomark. *16*, 161-169.

Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., Sun, W., Dou, K., and Li, H. (2015a). Circular RNA: a new star of noncoding RNAs. Cancer Lett. *365*, 141-148.

Qu, S., Song, W., Yang, X., Wang, J., Zhang, R., Zhang, Z., Zhang, H., and Li, H. (2015b). Microarray expression profile of circular RNAs in human pancreatic ductal adenocarcinoma. Genom. Data *5*, 385-387.

Ravn, U., Didelot, G., Venet, S., Ng, K.-T., Gueneau, F., Rousseau, F., Calloud, S., Kosco-Vilbois, M., and Fischer, N. (2013). Deep sequencing of phage display libraries to support antibody discovery. Methods *60*, 99-110.

Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., and Ashwal-Fluss, R. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. Mol. Cell *58*, 870-885.

Salzman, J. (2016). Circular RNA expression: its potential regulation and function. Trends Genet. *32*, 309-316.
Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. PLoS Genet. *9*, e1003777.

Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PloS One *7*, e30733.

Sand, M., Bechara, F.G., Gambichler, T., Sand, D., Bromba, M., Hahn, S.A., Stockfleth, E., Hessam, S. (2016). Circular RNA expression in cutaneous squamous cell carcinoma. J. Dermatol. Sci. *83*, 210-218.

Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., and Ang, X.L. (2008). Cancer proliferation gene discovery through functional genomics. Science *319*, 620-624.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., and Doench, J.G. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84-87.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science *319*, 617-620.

Song, X., Zhang, N., Han, P., Moon, B.-S., Lai, R.K., Wang, K., and Lu, W. (2016). Circular RNA profile in gliomas revealed by identification tool UROBORUS. Nucleic Acids Res. *44*, e87-e87.

Strezoska, Ž., Licon, A., Haimes, J., Spayd, K.J., Patel, K.M., Sullivan, K., Jastrzebski, K., Simpson, K.J., Leake, D., and van Brabant Smith, A. (2012). Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens. PLoS One *7*, e42341.

Szabo, L., Morey, R., Palpant, N.J., Wang, P.L., Afari, N., Jiang, C., Parast, M.M., Murry, C.E., Laurent, L.C., and Salzman, J. (2015). Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. Genome Biol. *16*, 126.

Tolle, F., and Mayer, G. (2016). Preparation of SELEX Samples for next-generation sequencing. Methods Mol. Biol. *1380*, 77-84.

Van Blarcom, T., Rossi, A., Foletti, D., Sundar, P., Pitts, S., Bee, C., Witt, J.M., Melton, Z., Hasa-Moreno, A., and Shaughnessy, L. (2015). Precise and efficient antibody epitope determination through library design, yeast display and next-generation sequencing. J. Mol. Biol. *427*, 1513-1534.

Van Der Meer, R., Song, H.Y., Park, S.-H., Abdulkadir, S.A., and Roh, M. (2014). RNAi screen identifies a synthetic lethal interaction between PIM1 overexpression and PLK1 inhibition. Clin. Cancer Res. *20*, 3211-3221.

Vicens, Q., and Westhof, E. (2014). Biogenesis of circular RNAs. Cell *159*, 13-14.

Vizeacoumar, F.J., Arnold, R., Vizeacoumar, F.S., Chandrashekhar, M., Buzina, A., Young, J.T., Kwan, J.H., Sayad, A., Mero, P., and Lawo, S. (2013). A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. Mol. Syst. Biol. *9*, 696.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., and Perou, C.M. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. *38*, e178-e178.

Wang, P.L., Bao, Y., Yee, M.-C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O., and Salzman, J. (2014a). Circular RNA is expressed across the eukaryotic tree of life. PloS One *9*, e90859.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014b). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80-84.

Wang, X., Zhang, Y., Huang, L., Zhang, J., Pan, F., Li, B., Yan, Y., Jia, B., Liu, H., and Li, S. (2015). Decreased expression of hsa_circ_001988 in colorectal cancer and its clinical significances. Int. J. Clin. Exp. Pathol. *8*, 16020–16025.

Weng, W., Wei, Q., Toden, S., Yoshida, K., Nagasaka, T., Fujiwara, T., Cai, S., Qin, H., Ma, Y., and Goel, A. (2017). Circular RNA ciRS-7—a promising prognostic biomarker and a potential therapeutic target in colorectal cancer. Clin. Cancer Res. *23*, 3918-3928.

Wilusz, J.E., and Sharp, P.A. (2013). A circuitous route to noncoding RNA. Science *340*, 440-441.

Xia, S., Feng, J., Chen, K., Ma, Y., Gong, J., Cai, F., Jin, Y., Gao, Y., Xia, L., and Chang, H. (2017). CSCD: a database for cancer-specific circular RNAs. Nucleic Acids Res. *46*, D925-D929.

Xuan, L., Qu, L., Zhou, H., Wang, P., Yu, H., Wu, T., Wang, X., Li, Q., Tian, L., and Liu, M. (2016). Circular RNA: a novel biomarker for progressive laryngeal cancer. Am. J. Transl. Res. *8*, 932-939.

Xue, Q., Sun, K., Deng, H.-J., Lei, S.-T., Dong, J.-Q., and Li, G.-X. (2013). MicroRNA-338-3p inhibits colorectal carcinoma cell invasion and migration by targeting smoothened. Jpn. J. Clin. Oncol. *44*, 13-21.

Yang, Q., Du, W.W., Wu, N., Yang, W., Awan, F.M., Fang, L., Ma, J., Li, X., Zeng, Y., and Yang, Z. (2017a). A circular RNA promotes tumorigenesis by inducing c-myc nuclear translocation. Cell Death Differ. *24*, 1609-1620.

Yang, W., Du, W., Li, X., Yee, A., and Yang, B. (2016). Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis. Oncogene *35*, 3919-3931.

Yang, Z.-G., Awan, F.M., Du, W.W., Zeng, Y., Lyu, J., Wu, D., Gupta, S., Yang, W., and Yang, B.B. (2017b). The circular RNA interacts with STAT3, increasing its nuclear translocation and wound repair by modulating Dnmt3a and mir-17 function. Mol. Ther. *25*, 2062-2074.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., and Quedenau, C. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. Nat. Neuros. *18*, 603-610.

Zhang, H.D. Jiang, L.H., Sun, D.W., Hou, J.C., Ji, Z.L. (2017). CircRNA: a novel type of biomarker for cancer. Breast Cancer, *25*, 1-7.

Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. Cell *159*, 134-147.

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., and Chen, L.-L. (2013). Circular intronic long noncoding RNAs. Mol. Cell *51*, 792-806.

Zheng, Q., Bao, C., Guo, W., Li, S., Chen, J., Chen, B., Luo, Y., Lyu, D., Li, Y., and Shi, G. (2016). Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nat. Commun. *7*, 11215.

Zhong, Z., Lv, M., and Chen, J. (2016). Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. Sci. Rep. *6*, 30919.