

# Investigation of the roles of circular RNA circZNF609 during colorectal cancer progression



## **DISSERTATION**

zur Erlangung des  
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)  
der Fakultät Biologie und Vorklinische Medizin  
der Universität Regensburg

Vorgelegt von  
**Hùng Hồ Xuân**  
aus Nghe An, Vietnam

Regensburg, 2017



Das Promotionsgesuch wurde eingereicht am:

13.12.2017

Die Arbeit wurde angeleitet von:

Prof. Dr. Gunter Meister

---

Hung Ho Xuan



*For my family*

*“Two roads diverged in a wood, and I—*

*I took the one less traveled by,*

*And that has made all the difference.”*

*Robert Frost, 1874-1963.*



# Table of Contents

<b>SUMMARY .....</b>	<b>11</b>
<b>1 INTRODUCTION .....</b>	<b>12</b>
1.1 Discovery of circRNAs .....	12
1.2 CircRNA properties.....	13
1.3 CircRNA biogenesis .....	13
1.3.1 CircRNA biogenesis regulated by cis elements .....	14
1.3.2 CircRNA biogenesis regulated by RBPs .....	16
1.4 CircRNA functions .....	19
1.4.1 CircRNAs can function as miRNA sponges.....	19
1.4.2 CircRNAs can serve as templates for translation.....	20
1.4.3 CircRNA associated with RBPs (circRBPs).....	23
1.4.4 Other functions of circRNAs .....	25
1.5 CircRNAs in development and diseases .....	26
1.6 Study about circZNF609 and its parental ZNF609.....	26
1.7 Colorectal cancer progression and metastasis.....	27
1.8 Aim of this thesis .....	29
<b>2 RESULTS .....</b>	<b>30</b>
2.1 Mouse model and RNAseq analysis for detection of circRNAs .....	30
2.2 Characterization of circZNF609 .....	31
2.3 Generation of shRNA constructs targeting circZNF609.....	33
2.4 Methods for overexpression of circZNF609.....	37

<b>2.5</b>	<b>CircZNF609 potentially encodes for proteins .....</b>	<b>38</b>
2.5.1	Flag reporter constructs and mutagenesis .....	38
2.5.2	Detection of circZfp609 proteins by anti-Zfp609 antibodies .....	48
<b>2.6</b>	<b>Synthesis of circZNF609 <i>in vitro</i> .....</b>	<b>51</b>
2.6.1	PIE system for <i>in vitro</i> transcription (IVT) of circRNAs .....	51
2.6.2	IVT for circZNF609 synthesis .....	52
2.6.3	Northern blot validation of synthetic circZNF609.....	53
2.6.4	<i>In vitro</i> translation of synthetic circZNF609 .....	55
2.6.5	Transfection of synthetic circZNF609 to mammalian cells .....	55
<b>2.7</b>	<b>Toward understanding of circZNF609 protein function .....</b>	<b>56</b>
2.7.1	Generation of stable cell lines .....	56
2.7.2	Subcellular localization of circZNF609 proteins.....	58
2.7.3	Identification of circZNF609 interaction partners.....	58
<b>2.8</b>	<b><i>In vitro</i> and <i>in vivo</i> validation the function of circZNF609 during colorectal cancer progression .....</b>	<b>60</b>
2.8.1	<i>In vitro</i> proliferation assay .....	60
2.8.2	<i>In vivo</i> xenograft mouse model .....	62
<b>3</b>	<b>DISCUSSION .....</b>	<b>66</b>
<b>3.1</b>	<b>Overexpression of circZfp609.....</b>	<b>66</b>
<b>3.2</b>	<b>Translation of circZfp609.....</b>	<b>67</b>
<b>3.3</b>	<b>CircRNA synthesis <i>in vitro</i>.....</b>	<b>70</b>
<b>3.4</b>	<b>Transfection of circRNAs into mammalian cells.....</b>	<b>71</b>
<b>3.5</b>	<b>Effect of circZNF609 on colorectal cancer progression <i>in vitro</i> and <i>in vivo</i>.....</b>	<b>72</b>
<b>3.6</b>	<b>Potential function of circZNF609 as miRNA sponges .....</b>	<b>73</b>
<b>3.7</b>	<b>Analysis of proteins generated from circZNF609 .....</b>	<b>73</b>
<b>3.8</b>	<b>Interaction partners of circZNF609 proteins.....</b>	<b>74</b>



<b>4</b>	<b>OUTLOOK.....</b>	<b>76</b>
<b>5</b>	<b>APPENDIX .....</b>	<b>77</b>
<b>6</b>	<b>MATERIALS AND METHODS.....</b>	<b>107</b>
<b>6.1</b>	<b>Materials.....</b>	<b>107</b>
6.1.1	Instruments .....	107
6.1.2	Plasmids .....	108
6.1.3	Oligos .....	109
6.1.4	Antibodies.....	114
6.1.5	Bacterial strains.....	115
6.1.6	Mammalian cell lines .....	115
6.1.7	Mouse strains .....	115
6.1.8	Buffers and solutions .....	115
<b>6.2</b>	<b>Methods.....</b>	<b>120</b>
6.2.1	Methods for working with DNA.....	120
6.2.2	Methods for working with RNA.....	124
6.2.3	Methods for working with proteins.....	128
6.2.4	Methods for working with cells .....	131
6.2.5	Methods for working with animals .....	134
6.2.6	Next generation sequencing and data analysis .....	136
<b>7</b>	<b>REFERENCES.....</b>	<b>137</b>
<b>8</b>	<b>CURRICULUM VITAE .....</b>	<b>151</b>
<b>9</b>	<b>PUBLICATIONS AND PRESENTATIONS .....</b>	<b>153</b>
<b>9.1</b>	<b>Publications.....</b>	<b>153</b>
<b>9.2</b>	<b>Presentations .....</b>	<b>153</b>
<b>10</b>	<b>ACKNOWLEDGEMENTS .....</b>	<b>155</b>



## Summary

Non-coding RNAs can function as potent gene regulators and consequently, such RNA molecules have been implicated in many diseases including different types of cancer. Circular RNAs (circRNAs) are emerging as a new class of regulatory non-coding RNAs. Several studies have revealed that circRNAs are differentially expressed in several tissues and thus, some circRNAs were reported to play important roles in several diseases such as cancer, cardiovascular disease and neurological disease.

Using an intrasplenic tumor model for colorectal cancer, we have identified several circRNAs differentially expressed in primary tumors and liver metastases suggesting a role of these RNAs in tumor progression. In this study, we have focused on characterization of circular RNA ZNF609 (circZNF609), which is up-regulated in primary tumors as well as liver metastases in our mouse model. We developed methods for knockdown and overexpression of circZNF609 and established several inducible stable cell lines for *in vitro* and *in vivo* analyses. Using two xenograft mouse models, we confirmed that circZNF609 can function as oncogene to promote colorectal cancer progression.

Interestingly, we found that circZNF609 encodes for small proteins, which could potentially have function in tumor development. Using Flag reporter constructs and mutagenesis, we identified several AUGs (even exogenous AUGs) which can be used to initiate translation from circZNF609. Moreover, circZNF609 protein localizes to the cytoplasm while the parental ZNF609 protein exclusively localizes to the nucleus. Mechanistically, mass spectrometry analysis identified several proteins of the integrator complex as interaction partners of both circZNF609 and ZNF609 proteins suggesting an interesting mode of circZNF609 functions. Further analysis of the molecular mechanisms of these interactions would shed lights on how circZNF609 functions.

# 1 Introduction

## 1.1 Discovery of circRNAs

CircRNAs are emerging as a new class of regulatory RNAs among several well-studied non-coding RNAs (ncRNAs) including small non-coding RNAs, long non-coding RNAs, or enhancer RNAs. Yet, circRNAs were discovered several years ago, first in the late 70s from plant viroids<sup>1,2</sup> and then in the 80s from yeast mitochondrial RNAs<sup>3,4</sup> intervening sequence of yeast ribosomal RNA precursor<sup>5,6,7</sup> and hepatitis delta virus<sup>8</sup>. Subsequently, circRNAs were reported in mammalian cells as scrambled transcripts<sup>9,10</sup> and considered as by-products of canonical splicing due to its low expression levels<sup>11</sup>. However, in some cases, circRNAs were found to be very abundant, such as Sry circRNA in mouse testis<sup>12</sup> or NCX1 circRNA<sup>13</sup> and therefore suggested as functional products of canonical splicing<sup>12</sup>. Such circular events were more and more observed in the following years in several species and genes including the human and rat cytochrome <sup>14,15,16</sup>; human dystrophin<sup>17</sup>, human and blood ETS-1 <sup>18</sup> or rat, mouse, rabbit and monkey NCX1 <sup>13</sup>. Moreover, circular RNAs were successfully recapitulated *in vitro* using the ATC1 derived pre-mRNA transcript in yeast either from pseudocircular pre-mRNA<sup>19</sup> or linear pre-mRNA<sup>20</sup> and later on using the human  $\beta$ -globin intron<sup>21</sup> or PIP71 and PIP73 exons flanking by short intronic sequences<sup>22</sup> in Hela cell extracts. These evidences confirm the existence of circRNAs but to identify those transcripts systematically was not possible at that time due to the lack suitable high-throughput technologies.

An attempt to identify those circularized exons in 2012 in a high throughput manner led to the discovery of thousands of circRNA transcripts<sup>23</sup>. This is the first time, a systematic verification of circRNAs was confirmed showing that circRNAs are non-polyadenylated, predominantly cytoplasmic-localized and expressed at somehow similar level compared to their canonical linear isoforms in leukocytes. Subsequently, advance in next-generation RNA sequencing (RNA-Seq) of nonpolyadenylated or rRNA-depleted transcriptomes has recently updated the widespread expression of circRNAs in different organisms<sup>24,25,26</sup>, some of which have been biochemically confirmed to be highly expressed in a cell type-specific or tissue-specific manner<sup>27,25,28,29</sup>. Bioinformaticians have developed various algorithms to detect genome-wide circRNA expression from RNA-Seq data in the past few years and a source of experimental and bioinformatic approaches has been addressed and discussed in several reviews<sup>30,31,32,33,34</sup>.

## 1.2 CircRNA properties

CircRNAs are covalently closed single-stranded transcripts and resistant to the degradation by exonucleases, such as RNase R<sup>35,36</sup>. Since circRNAs do not contain a 5' cap and a 3' tail, they are very stable with half-lives much longer than observed for linear RNA counterparts<sup>37,28,38</sup>.

Depending on how circRNAs are generated, 3 types of circRNAs are observed in cells across several species: exonic circRNAs, intronic circRNAs, and exon-intron circRNAs<sup>39</sup> (Fig. 1.1 A). Exonic circular RNAs are the most abundant and well-studied circRNAs<sup>40</sup>. These circRNAs are very conserved between different species and localized predominantly in the cytoplasm<sup>28,38</sup>. Intronic circRNAs were first discovered in the early 80s, many of those came from intron lariats and resist to debranching enzymes<sup>41,42</sup>. Exon-intron circRNAs, or ElciRNAs, come from back-splicing and are regulated by complementary sequence flanking its circularized exon-intron part<sup>43</sup>. In general, intron-containing circRNAs localizes exclusively in the nucleus<sup>42,43</sup>.

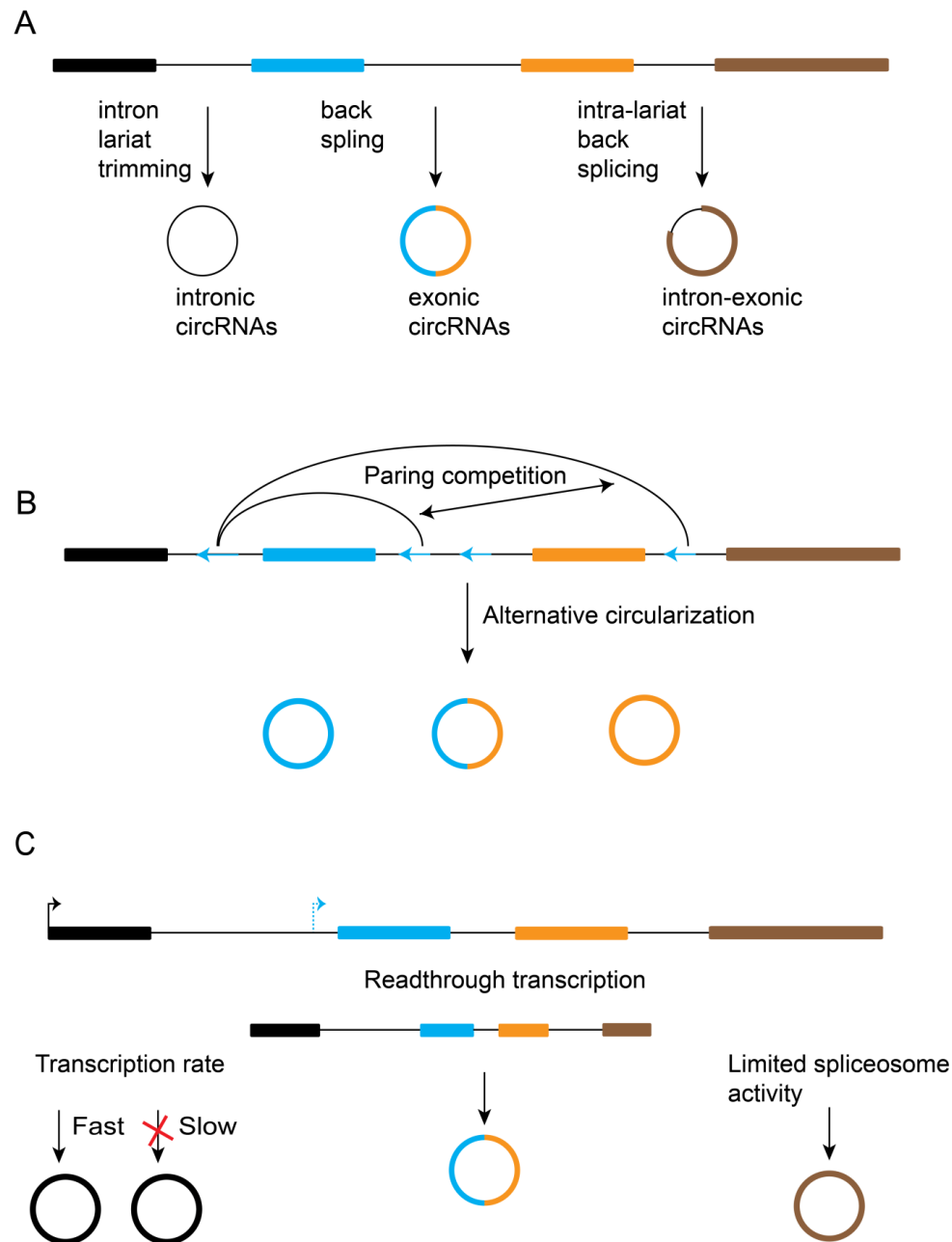
## 1.3 CircRNA biogenesis

CircRNAs were initially considered as by-products of aberrant splicing<sup>10,11</sup>. Recent studies about circRNA biogenesis have revealed that circRNAs are produced via back-splicing, a highly regulated process controlled by several cis and trans regulators<sup>44–50,51,52,53</sup>. In contrast to the canonical splicing, back-splicing joins a downstream splice donor site with an upstream splice acceptor site<sup>46</sup> (Fig. 1.1 A). However, how the spliceosome and which factors of the spliceosome are involved in back-splicing is not clear<sup>54</sup>. Mutagenesis of splice sites and using splicing inhibitor isoginkgetin to block spliceosome assembly showed that canonical spliceosomal machinery is required for back-splicing<sup>53</sup>. Moreover, it was proposed that linear splicing and back-splicing may compete with each other, suggesting that both processes share some similar splicing activities<sup>45</sup> while recent research also showed that regulation of back-splicing has different rules and therefore requires different splicing factors<sup>55</sup>. Indeed, very recently, it was shown that circRNA expression increases while expression of their cognate linear mRNAs decreases when the core spliceosomal components are depleted<sup>56</sup>. Nevertheless, it is now clear that back-splicing is promoted by regulatory elements residing in the flanking introns of the circularized exons<sup>44,51,52,49</sup> and could be repressed or enhanced by RNA-binding proteins (RBPs)<sup>50,47,52,53,55,57,58,59,60</sup>. Still, it is not clear how back-splicing is linked to transcription. It

is still a matter of debate of whether circRNAs are produced co-transcriptionally<sup>48,51</sup> or post-transcriptionally<sup>45,52</sup>. It was also reported that circular RNAs from downstream genes could be produced via readthrough transcription<sup>56</sup> (Fig. 1.1 C).

### 1.3.1 CircRNA biogenesis regulated by cis elements

In the late 90s, circular RNAs were successfully recapitulated *in vitro*<sup>19–22</sup> and many evidences implied that the complementary sequence of flanking intron for base pairing between 5'- and 3'-ends is required for circRNA production<sup>20,12,61</sup>. Genome-wide analysis to identify cis-sequence elements using sequences upstream or downstream from backsplice sites revealed the presence of Alu repeats in flanking introns, which are often very long<sup>28,49</sup>. Since Alu repeats are only presented in primates, analysis of circRNAs in other organisms is necessary to understand circRNA biogenesis mechanism. Analysis of introns from *Caenorhabditis elegans* indicated reverse complementary matches are important for circRNA biogenesis<sup>44</sup>. Therefore, reverse complementary sequence, including Alu elements, are now considered as a conserved feature of circRNA biogenesis since the pairing of those sequence to form RNA duplexes significantly enhances back-splicing<sup>49,51,52</sup>. Indeed, those reverse complementary sequences have been validated experimentally and are becoming useful tools for circRNA overexpression<sup>51,52</sup> or circRNA knockout via CRISPR/Cas9<sup>62</sup>. Interestingly, due to the availability of several reverse-orientated Alu elements in one intron, competition of these elements for RNA pairing across intron can lead to the formation of different circRNAs. This process is called alternative circularization<sup>49</sup>. This event together with alternative splice site selection leads to diversity in the landscape of circRNAs<sup>63,64</sup> (Fig. 1.1 B).



**Fig. 1.1 Biogenesis of circRNAs** A) Three types of circRNAs were detected and profiled from cells. Intronic circRNAs are generated from intron lariats and are resistant to debranching enzymes<sup>42</sup>. Exonic circRNAs are generated via back-splicing of one or several exons<sup>40</sup>. Exon-intron circRNAs are circularized exons with part of retained flanking intron<sup>43</sup>. B) Alternative circularization mechanisms. Multiple circRNAs can be produced from a single gene locus with different numbers of exons included. This is regulated by the competition of RNA pairing across different introns<sup>49</sup>. C) CircRNAs can be regulated through transcription. For examples, in neurons, circRNAs are upregulated because of fast transcription and accumulation<sup>48</sup>. In some cases, production of circular RNAs from downstream genes can be generated through back-splicing of a readthrough transcripts produced from the promoter of upstream gene<sup>56</sup>. In other cases, circRNAs are upregulated when the core components of spliceosomes are depleted<sup>56</sup>.

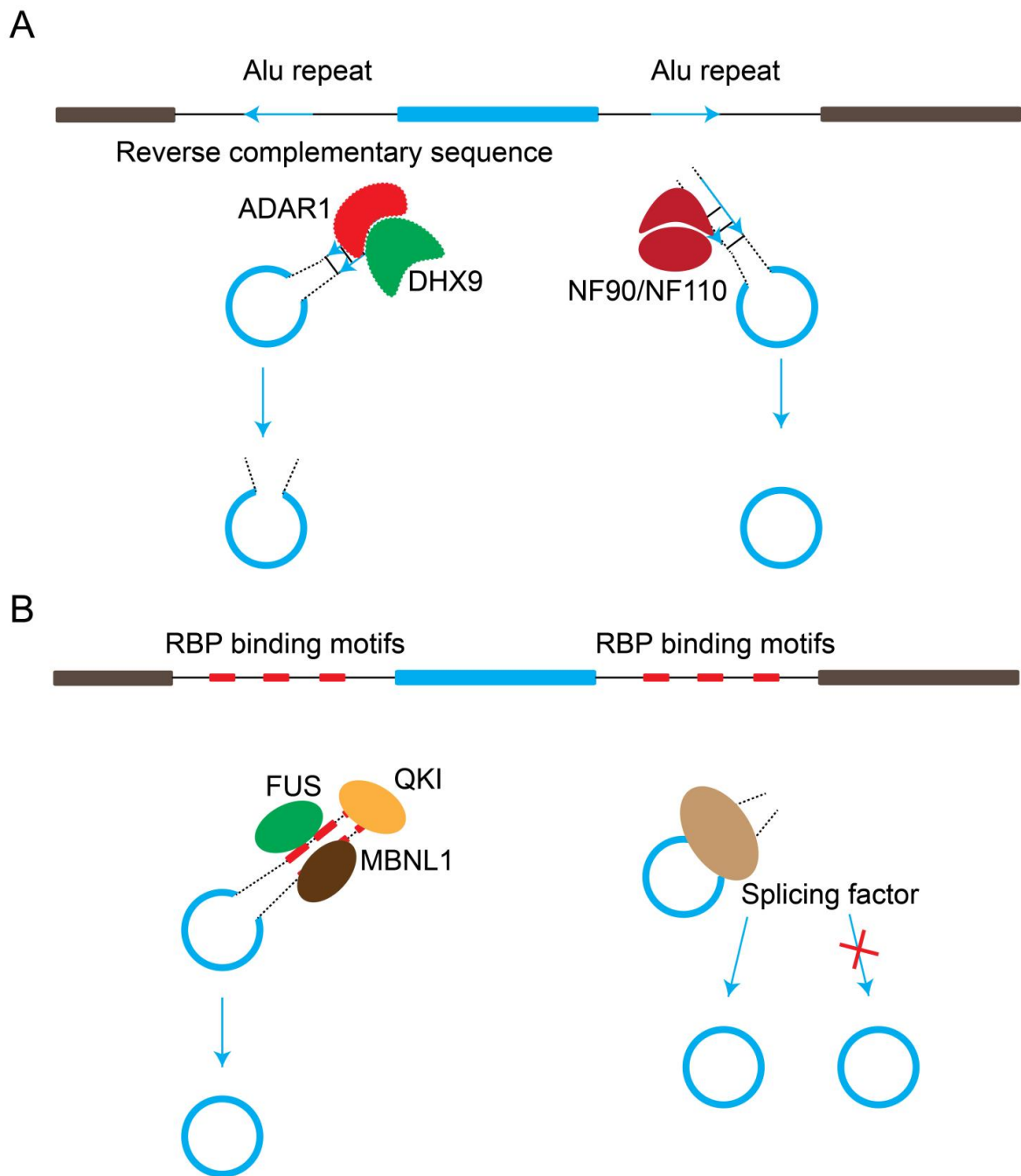
### 1.3.2 CircRNA biogenesis regulated by RBPs

Recent works indicated that back-splicing activities are not only driven by cis intron-pairing but can also be affected by trans-acting factors such as RBPs. In some cases, the production of circRNAs is controlled by several proteins acting in a combinatorial manner, for instance, the action of multiple heterogeneous nuclear ribonucleoprotein hnRNP and serine–arginine SR proteins<sup>52</sup>. In other cases, the production of several circRNAs, for instance, 80 different circRNA isoforms of the titin gene, are regulated by a single protein as shown for the splicing factor RBM20<sup>60</sup>. Indeed, ADAR1<sup>44,65</sup> and DHX9<sup>57</sup> antagonize circRNA formation while QKI<sup>50</sup>, MBL/MBNL1<sup>45</sup>, hnRNPs and SR protein<sup>52</sup> and NF90/NF110<sup>47</sup> promote circularization. Moreover, HNRNPL<sup>59</sup> or FUS<sup>58</sup> could either enhance or inhibit circRNA production (Fig. 1.2).

Analysis of flanking introns of circRNAs also revealed that these sequences are enriched in adenosine to inosine (A-to-I) editing and therefore ADAR1 was proposed to binds to these pairing intron regions and converts adenosines to inosine<sup>44</sup>. This was experimentally confirmed by the fact that the production of a subset of circRNAs was upregulated when ADAR1 is depleted in different human and mouse cell lines<sup>65,44</sup>. Another study showed that DHX9, a nuclear RNA helicase, binds specifically to inverted-repeat Alu as well as interferon-inducible isoform of ADAR (p150) in an RNA-independent interaction manner. This indicates that DHX9 may antagonize circRNA formation. Indeed, loss of DHX9 leads to an increase of several circRNAs and co-depletion of ADAR and DHX9 enhances circular RNA production, suggesting a cooperative actions between these two enzymes in circRNA biogenesis<sup>57</sup> (Fig. 1.2 A).

A screen for splicing factors that regulate circRNA biogenesis using a dual color reporter construct, circScreen, showed that the production of several abundant circRNAs is regulated by QKI, an alternative splicing factor<sup>50</sup>. QKI binds to its consensus motif within introns flanking the circRNA-forming exon. Indeed, insertion of QKI motifs is sufficient to induce novel circRNA formation. Moreover, QKI knockdown or overexpression showed downregulation or upregulation of several circRNAs, therefore confirming its roles in regulating circRNA expression<sup>50</sup> (Fig. 1.2 B).





**Fig. 1.2 Factors that regulate circRNA biogenesis.** A) Exons of circRNAs are usually located in between very long introns<sup>28</sup>. These intron sequences contain Alu elements<sup>28</sup> or reverse complementary sequence<sup>44</sup> that forms RNA duplexes. The RNA pairing between the flanking introns facilitates circRNA production<sup>44,49</sup>. Some RBPs such as ADAR or DHX9 bind to these RNA duplexes and break the pairing between these RNA duplexes<sup>44,57</sup>. Some other RBPs such as NF90/NF110 bind and enhance RNA pairing between Alus, and therefore promote circRNA production<sup>47</sup>. B) Some RBPs binds to their own motifs in the flanking intron. The binding of these RBPs such as FUS<sup>58</sup>, QKI<sup>50</sup> or MBNL1<sup>45</sup> enhances circRNA production. In other cases, the production of circRNAs could be positively or negatively regulated by splicing factor such as HNRNPL<sup>59</sup> or spliceosomal components<sup>56</sup>.

Another attempt using genome-wide siRNA screening coupled to dual color reporter strategies identified several RBPs which could regulate circRNA biogenesis<sup>47</sup>. Specifically, the dsRNA-binding proteins NF90/NF110 bind RNA duplexes formed by inverted repeat elements in the introns flanking the circularized exon. Further analysis revealed that those inverted repeat elements are intronic Alu repeat and i-CLIP indicated that NF90 preferred to bind clusters of A-rich or U-rich sequences which are located to Alus in introns. This binding was shown to stabilize pairing introns for production of circRNAs (Fig. 1.2 B). Some circRNAs were shown to be downregulated upon NF90 depletion and the expression of those circRNAs was rescued when Flag-NF90, but not Flag-NF90 truncations, is re-expressed in those knockdown cells<sup>47</sup>. Upon poly(I:C) treatment or VSV infection, it was suggested that NF90/110 translocate to the cytoplasm and therefore downregulate the expression of nascent circRNAs. In untreated cells, RIP experiments showed that NF90/110 binds to some circRNAs but these binding reduced upon treatment. Moreover, vsvm mRNAs compete with circDHX34, but not linear DHX34 mRNA, for binding to NF90. This suggested a model that upon viral infection, the expression of circRNAs are downregulated due to translocation of NF90/NF110 to the cytoplasm, and there NF90/NF110 can bind to viral mRNAs for immune response<sup>47</sup>.

Apart from binding to the Alu repeat sequence, several RBPs were shown to bind to its own motif in the introns juxtaposing the circRNA-forming exon, as shown from QKI. circRNA transcriptome profiling for FUS WT/KO cells<sup>58</sup> or HNRNPL WT/KD cells<sup>59</sup> identified several up- or down-regulated circRNAs. Specifically, up-regulated circRNAs significantly associate with HNRNPL binding to its own motifs and down-regulated circRNAs to some extent changed upon HNRNPL knockdown. Interestingly, minigene constructs containing the HNRNPL binding site (CA)<sub>20</sub> to both of the flanking introns significantly enhances circRNA formation<sup>59</sup>. Another study showed that FUS depletion and mutation affect circRNA expression and FUS interacts with its binding sites in pre-mRNA regions controlling biogenesis of a few specific circRNAs<sup>58</sup> (Fig. 1.2 B).

Another interesting example comes from the splicing factor Muscleblind (Mbl) *Drosophila melanogaster* or MBNL1 in human<sup>45</sup>. Mbl regulates circularization of its own exon 2 from its pre-mRNA by binding to the putative Mbl binding site in the flanking intron. This was further proven by using minigene containing Mbl binding sites in the circMbl flanking introns under Mbl overexpression (Fig. 1.2 B). Specifically, Mbl overexpression augments the production of circMbl or circTim flanked by Mbl introns<sup>45</sup>. Furthermore, circMbl was

shown to bind to its parental protein, Mbl, and control expression of its own expression. It was proposed that when the expression of Mbl protein is high, the expression of Mbl mRNA is decreased by enhancing the production of circMbl, which then sponges excessed MBL protein<sup>45</sup>.

## 1.4 CircRNA functions

### 1.4.1 CircRNAs can function as miRNA sponges

The competing endogenous RNA (ceRNAs) hypothesis proposes that specific RNAs can sequester miRNAs and impair miRNA activity and therefore upregulate mRNA target gene regulation<sup>66,67</sup>. circRNAs are presumably ideal candidates for miRNA sponges (Fig. 1.3 F) because they are more stable and they lack both 5' and 3' ends, which facilitate escaping from exonuclease and degradation<sup>68,69</sup>. Indeed, the circRNA CDR1as (or ciRS-7) contains more than 70 miR-7 seed sequences and associates with Ago2 without being degraded<sup>70,26</sup>. Interestingly, knockout of CDR1as in mice shows miR-7 deregulation and increases the expression of targets of miR-7<sup>71</sup>. The second circRNA namely Sry contains 16 binding sites for miR-138 and therefore is proposed to act as a sponge<sup>70</sup>. However, only CDR1as and Sry stood out as two exceptional cases because such multiple binding sites for miRNA in one specific transcript is usually not found in nature. In fact, a comprehensive analysis also pointed out that most circRNAs do not contain several miRNA binding sites, except a few circRNAs from the ZNF gene family<sup>69</sup>. Moreover, given that some circRNAs are lowly expressed, a change in expression of a circRNA would not be sufficient to titrate miRNA away from their target<sup>69</sup>. This finding is supported by earlier study using high-throughput assay called Sensor-seq for quantification the activity of hundreds of miRNAs. This work showed that only the most abundant miRNAs in a cell can mediate target suppression<sup>72</sup>. However, it was also showed that miRNA-target ratios determine the potential for ceRNA regulation<sup>73</sup>. Specifically, highly expressed miR-294 and let-7 families are unsusceptible to ceRNAs while active miR-25 family, which has lower miRNA:target ratios, may be more sensitive to ceRNA activities<sup>73</sup>. Since ceRNA hypothesis has been actively debated<sup>67</sup>, any experiments of circRNA-miRNA sponges should be carefully performed to get a convincing conclusion.

Moreover, binding of miRNAs and Ago2 to a circRNA does not necessarily lead to miRNA suppression. For instance, ciRS-7 contains only one binding site for miR-671 and is cleaved by Ago2 in a RISC-dependent manner<sup>74,70</sup>. Therefore, binding of miRNAs and

Ago2 may have different functions than miRNA sponges. Depending on the sequence context of binding between circRNAs and miRNAs, the association of miRNA and Ago2 to circRNA might lead to circRNA degradation. Still, several studies suggested that miRNA sponges as a common feature of several circRNAs (Table 2).

#### 1.4.2 CircRNAs can serve as templates for translation

circRNA translation is not new. Before the era of long non-coding RNAs, every circRNA found was asked for coding potential by recapitulating its ORF over circRNA junction. The first circRNA with coding potential was discovered in 1985 and came from an intron circRNA of archaeobacterium *Desulfurococcus mobilis*<sup>75</sup>. This 622 bp circRNA is proposed to generate a protein of 194 aa with a putative Shine-Dagano sequence, GGGGU, that is complementary to 3' end of 16S RNA<sup>76</sup>. This circRNA was then proved to encode for a protein called I-Dmo I which functions as an endonuclease<sup>77</sup>. Afterwards, 2 other intronic circRNAs from the hyperthermophilic archaeon *Pyrobaculum organotrophum* were also proposed to encode for proteins<sup>78</sup>.

In mammals, the first circRNA was mentioned to have coding potential and its protein was suggested to be functional is from an abundant circRNA produced from NCX1 gene<sup>13</sup>. This is consistent with the observation that circRNA with internal ribosome entry sites (IRES) could be translated *in vitro*<sup>79</sup>. Notably, ribosome profiling, or Ribo-seq, is a powerful method for analyzing translation *in vivo* using next generation sequencing of ribosome-protected mRNA fragments<sup>80,81</sup>. Therefore, this could be also applied for studying circRNA translation if its RNA fragments are protected by ribosomes. However, only fragment spanning junction of circRNAs can distinguish translation of circRNA or mRNA counterpart. Due to the fact that most of the circRNAs do not associate with ribosomes<sup>69,82,83,28</sup> and the difficulties in identifying the junction reads of circRNAs from Ribo-seq data, circRNAs are considered as a new class of non-coding RNAs. Nevertheless, few exceptions are recently reported<sup>84–86,87</sup> and a protein product from circFBXW7 was indeed shown to have function in glioma tumorigenesis<sup>88</sup>.

Although circRNAs do not contain a 5' cap and a 3' poly-A tail, they form a circular configuration which may favor translation of their embedded ORFs (Fig. 1.3 D). Since circRNAs do not contain a 5' 7-methylguanosine (m<sup>7</sup>G) cap, translation of circRNAs will supposedly follow cap-independent initiation. Especially, during stress, while cap-dependent is repressed<sup>89,90</sup>, cap-independent translation is favored from the cells<sup>91</sup>. For

examples, during heat shock, cap-independent translation can be driven by N6-Methyladenosine (m6A)<sup>92,93</sup> besides of the use of IRES<sup>90</sup>.

It was shown that prokaryotic ribosomes can bind to circRNAs and eukaryotic ribosomes can initiate translation of circRNA *in vitro*<sup>79</sup>. Later on, a circRNA construct of *in vitro* transcribed GFP ORF, which contained a Shine-Dalgarno sequence, was shown to be translated *in vivo* in *E. coli*<sup>94</sup>. This construct produced 30 kDa GFP proteins from 795 nts circular mRNA transcript. The author then generated an infinite ORF in the reading frame of the initiating AUG by removing its in-frame stop codon. Western blot from *E. coli* strains containing this infinite GFP constructs showed several protein bands ranging in size from 50 to more than 300 kDa. Therefore, it was suggested that ribosomes can “travel” on the circRNA for several rounds, at least in the infinite GFP construct<sup>94</sup>. Such rolling circle translation was reported in prokaryotes<sup>95</sup>, in circular virusoid of the rice yellow mottle virus (sobemovirus)<sup>85</sup> as well as in living human cells<sup>96</sup>. It was reported that 220-nts CCC RNA from virusoid use the initiation–termination sequence UGAUGA to initiate translation. The ribosomes read in two (or three due to read-through mechanism) overlapped reading frames leading to new amino acid sequences. Interestingly, no IRES sequences was found in front of the AUG or after the termination codon of this circRNA<sup>85</sup>. A similar observation was reported later using infinite Flag construct. This study showed that rolling circle translation happens without any specific features<sup>96</sup>, such as IRES sequences. This also confirmed the suggestion that ribosomes can translate a circRNA several rounds<sup>94,96</sup>.

Several artificial minigene constructs were used to demonstrate that circRNA can be translated *in vivo* in human cells with or without a functional IRES element<sup>55,52,96</sup>. Other efforts to search for evidence of translation of endogenous circRNAs in mammalian cells showed that translation of circRNAs occurs in a very sophisticated and distinct manner depending on different circRNAs<sup>84,86,87,97</sup>. Indeed, translation of circRNAs can be driven by m6A<sup>84</sup>, by IRES<sup>87</sup> or coupled with splicing<sup>86</sup>, all of which using cap-independent mechanisms.

Firstly, since several circRNAs were shown to be m6A methylated<sup>98</sup>, circRNAs with m6A-containing short sequence were reported to be translated in cells<sup>84</sup>. This study showed that circRNA translation is promoted by METTL3/METTL14 and further requires eIF4G2 and YTHDF3. Furthermore, Mass spectrometry identified several endogenous peptides, which are generated from these m6A-containing circRNAs<sup>84</sup>.

Secondly, two studies suggested that circRNAs with IRES sequence can be translated *in vivo*. The first evidence came from circMbl<sup>87</sup>. This study showed that several circRNAs, including abundant circMbl, are associated with translating ribosomes in fly (and therefore referred as Ribo-circRNAs). Moreover, they also extended this finding by analyzing ribosome profiling data in murine C2C12 cells and rat brain and liver tissues. Interestingly, by analyzing such events, it was shown that Ribo-circRNAs are often not among the most abundant circRNAs and often share the 5' UTR and start codon with its linear parental mRNAs<sup>87</sup>. The UTR of circMbl can drive circRNA translation in a cap-independent manner and this translation is regulated by starvation and the RBP FOXO<sup>87</sup>. The second evidence of translation of circRNA driven by IRES came from circFBXW7. Translation of circFBXW7 is initiated via cap-independent manner and gives a novel 21 kDa protein, called FBXW7-188aa. To the best of my knowledge, this is the first study that showed a function of a protein generated from a circRNA. FBXW7-188aa suppresses glioma cell cycle and cell proliferation *in vitro* and *in vivo*. Specifically, FBXW7-188aa destabilizes c-Myc protein through binding to the de-ubiquitinating enzyme USP28. FBXW7-188aa shares N-terminal protein sequence with the E3 ligase FBXW7 $\alpha$ , the most abundant isoform of the FBXW7 gene. FBXW7 $\alpha$  binds to c-Myc via its C-terminal WD40 domain and binds to USP28 at its N-terminal domain. Given that USP28 stabilizes c-Myc, FBXW7-188aa competes with FBXW7 $\alpha$  for binding to USP28 and therefore liberates FBXW7 $\alpha$  for c-Myc degradation<sup>88</sup>.

Finally, another study showed that circRNA translation is dependent on splicing, as example of circZfp609 (or circZNF609 in human)<sup>86</sup>. Using an artificial construct for overexpression of circZfp609 with 3xFlag tag, circZfp609 was shown to generated 2 proteins from this minigene. The UTR of circZfp609 alone does not contain IRES activity as by viral IRES from ECMV but circZfp609 UTR rather needs splicing for function as IRES. Of notes, the construct circZfp609-3xFlag used in this study was generated by myself as part of the PhD project and provided to the Irene Bozzoni lab.

Such examples suggested us that translation from circRNAs may happen more frequently in nature and provide new sources of novel proteins regulating different aspects of biological activities. Many bioinformatics tools were developed to predict the availability of such translation events<sup>99,100,101,102</sup> and more translation studies of circRNAs will presumably be reported in the near future.

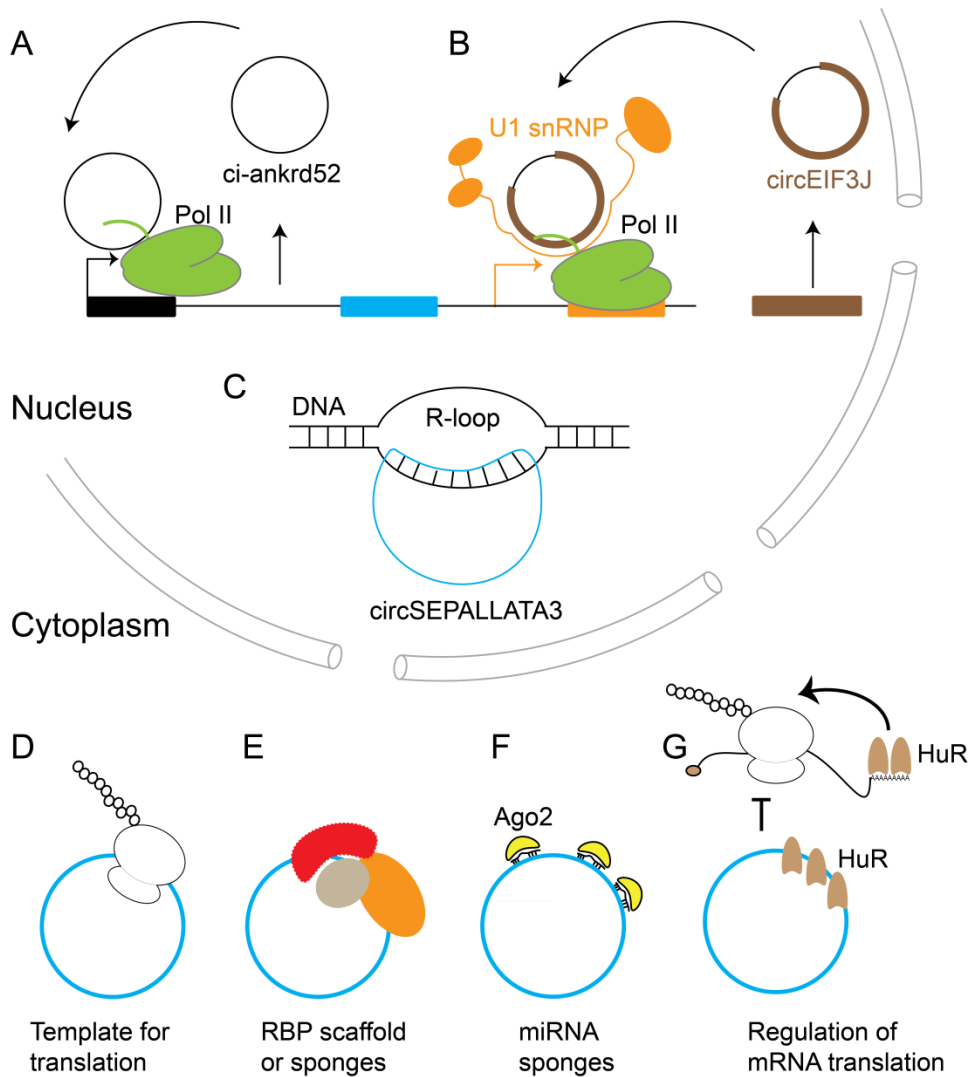
### 1.4.3 CircRNA associated with RBPs (circRBPs)

The development of new bioinformatic approaches revealed that certain RNA-binding proteins (RBPs) possess a high number of binding sites in circRNA sequences<sup>99</sup>. Several studies showed that circRNAs are specifically bound by RBPs and have different functions (Fig. 1.3 E).

One study showed that circFOXO3 represses cell cycle by forming a circFOXO3-p21-CDK2 complex, therefore blocking the function of CDK2 and arrest cell cycle progression<sup>105</sup>. Moreover, silencing of circFOXO3 inhibits senescence. Mechanistically, circFOXO3 interacts with ID-1 and E2F2 and anti-stress proteins FAK and HIF1 $\alpha$ <sup>107</sup>. Furthermore, circFOXO3 overexpression increase apoptosis through upregulation of FOXO3 and its downstream target PUMA. Mechanistically, circFOXO3 promotes MDM2-induced p53 ubiquitination, resulting in p53 degradation<sup>108</sup>. Another circRNA called circAmotl1 contains binding sites for c-Myc. Ectopic expression of circAmotl1 increased c-Myc retention in the nucleus and promoted c-Myc stability. Normally, c-Myc mainly localizes to the cytoplasm, translocating c-Myc to the nucleus promotes tumorigenesis<sup>109</sup>.

Other searches for circRBPs show that several circRNAs such as circCDYL, circNFATC3, and circANKRD17 were shown to bind to IMP3 proteins using co-IP coupled to RNA-Seq<sup>83</sup>. However, the functional consequence of the association of circRNAs with IMP3 is not known. Later, circANRIL was shown to bind to pescadillo homologue 1, or PES1, a component of a pre-ribosomal assembly complex<sup>110</sup>. circANRIL prevents rRNA maturation by competing for binding sites on PES1 with pre-rRNA. This leads to activation of p53 and induction of nucleolar stress and therefore increases cell apoptosis<sup>110</sup>. Such competitive binding was also shown for circPABPN1, which inhibits HuR binding to its cognate mRNA PABPN1. Since HuR is a translational activator, this competition leads to translation suppression of PABN1<sup>106</sup> (Fig. 1.3 G).

Other study found that overexpressed circZKSCAN1 binds to several RBPs including MOV10 or UPF1 and UPF2, components of the non-sense mediated decay pathway or YTFDF2, YTFDF3, proteins that read m6A modification<sup>111</sup>. Interesting, circGFP, which is not expressed in human cells but produced by circZKSCAN1 intron, was shown to bind to several RBPs, much more than circZKSCAN1 produced from its endogenous intron. However, the function of these interactions is not known and would be interesting to investigate in the future<sup>111</sup>.



**Fig. 1.3: Functions of circRNAs.** A) ci-ankrd52 is generated from second intron of the ANKRD52 gene and is escaped from debranching of its intron lariat. ci-ankrd52 accumulates in the nucleus and associates with RNA Pol II to regulate expression of its parental gene ANKRD52<sup>42</sup>. B) circEIF3J is generated from exon with retained intron of the EIF3J gene. It is also accumulated in nucleus and interacts with U1 snRNP and RNA Pol II to enhance the transcription of its own gene EIF3J<sup>43</sup>. C) circSEPALLATA3 (or SEP3) is derived from exon 6 of SEPALLATA3 mRNA. It forms RNA:DNA hybrid with its parental DNA to regulate the splicing of its cognate SEPALLATA3 mRNA<sup>103</sup>. D) Some circRNAs are associated with ribosomes and are proposed to function as template for translation<sup>86,87</sup>. E) Several circRNAs contains binding motifs for several RBPs and therefore are thought to function as RBP sponges or scaffold for RBPs<sup>104,105</sup>. F) Most of circRNAs are predominantly localized to the cytoplasm and contain binding sites for miRNAs, and therefore are suggested to function as miRNA sponges. For examples, CDR1as (or ciRS-7) contains more than 70 binding sites for sponging miR-7<sup>70,26</sup>. G) HuR binds to the 3' UTR of PABPN1 mRNA and enhances PABPN1 mRNA translation. Binding of HuR to circPABPN1 prevents binding of HuR and PABPN1 mRNA UTR and therefore decrease the translation of PABPN1 mRNA<sup>106</sup>.



The development of new bioinformatic approaches, such as a database called “CircInteractome”<sup>99</sup> or circScan<sup>100</sup>, allows to systematically search for potential interactions of circRNA with RBPs. As miRNAs sponges seem to be the common features of circRNAs, several circRNAs were shown to bind to Ago2. Consistently, CircInteractome and circScan identified several circRNAs-RPBs from CLIP-seq data, including Ago2<sup>99,100</sup>. Moreover, as several circRNAs contain m6A modifications<sup>98</sup>, circScan showed that YTHDFs are the proteins that bind to m6A-circRNAs<sup>100</sup>. Interestingly, some of these YTHDFs-bound circRNAs also bind to eIF3 and enrich in ribosome-bound fractions suggesting that some of these circRNAs can be translated<sup>100</sup>.

In the future, with the efforts in identifying the RBPs of circRNAs, we could understand more about the function of circRNAs, apart from miRNA sponges.

#### 1.4.4 Other functions of circRNAs

Apart from circRNAs localized to the cytoplasm, which function as miRNA or RBP sponges or associated with ribosomes, several circRNAs, mostly intron-containing circRNAs, were reported to localize to the nucleus<sup>42,43</sup>. Many circRNAs are associated with RNA polymerase II (Pol-II) polymerase<sup>43,42</sup> or form R-loops to regulate transcription of its parental genes<sup>103</sup>.

The first evidence came from bioinformatics analysis of poly (A) (-) transcriptomes. This study identified several circRNAs that come from intron lariat and can escape from debranching. One of those called ci-ankrd52 binds to the Pol-II machinery and positively enhances the transcription of its host ankrd52 pre-mRNA<sup>42</sup> (Fig. 1.3 A). Another study identified a class of exon-intron-containing circRNAs, called exon-intron circRNAs or ElciRNAs. Since ElciRNAs localize to the nucleus, it was suggested that the RNA-RNA interaction between ElciRNAs and U1 snRNAs forms the ElciRNAs-U1 snRNA complexes. This complex then interacts with Pol-II transcription machinery to promote gene expression (Fig. 1.3 B). For instance, circEIF3J and circPAIP2 were shown to regulate the expression of their parental genes<sup>43</sup>.

Another example came from circRNA from *Arabidopsis thaliana*. The circRNAs derived from exon 6 of the SEPALLATA3 gene called SEP3 promotes the expression of its cognate endogenous mRNA variant SEP3.3 which skips exon 6. Mechanistically, SEP3 circRNA binds strongly to its DNA locus, forming a RNA-DNA hybrid (Fig. 1.3 C). It is

suggested that this R-loop structure may stall the transcription elongation and therefore increases the production of exon-skipped transcript variant<sup>103</sup>.

## 1.5 CircRNAs in development and diseases

Advance in high-throughput technologies has profiled several circRNAs differently expressed during different developmental process, differentiation or cancer. A recent study published a resource of circRNA expression from 20 human tissues with disease-related clinical research<sup>29</sup> suggesting that circRNA can be future biomarkers for several diseases and potential implications for therapeutic and research applications. Together with other studies, it is now clear that the expression of circRNAs is highly regulated in cells and show tissue-specific expression patterns. For examples, circRNAs are most abundant in brain and regulate in neuronal plasticity<sup>82</sup>. Many circRNAs are involved in cardiovascular development and pathology as summarized by Li et al., 2017<sup>112</sup>. Several circRNAs are reported to play important roles in cancer progression and metastasis as summarized by Kristensen et al., 2017<sup>113</sup>. In most of the cases, the functions of circRNAs are related to miRNA sponges.

## 1.6 Study about circZNF609 and its parental ZNF609

The first study about zinc-finger protein ZNF609 is from its mouse ortholog, Zfp609<sup>114</sup>. It was shown that Zfp609 is necessary for Rag1 and Rag2 expression in developing thymocytes. Zfp608 blocks T-cell maturation by repressing the expression of Zfp609 and therefore represses the expression of Rag1 and Rag2<sup>114</sup>. Recently, Zfp609 was shown to bind to Nipbl and the Integrator complex and regulate genes that control neuronal migration<sup>115</sup>.

CircZfp609 was shown to be upregulated during primary neuron differentiation<sup>65</sup> and aging of the mouse brain<sup>116</sup>. Another study showed that the expression of human circZNF609 (was also called myocardial infarction-associated circular RNA or MICRA)<sup>117</sup> in blood samples is lower in myocardial infarction patients compared to healthy control samples. Moreover, the expression of MICRA in peripheral blood samples of patients with acute myocardial infarction could be used for prediction of left ventricular dysfunction. However, the mechanism of circZNF609 in these studies is not clear<sup>117</sup>.

Two other studies showed that circZNF609 has a function as miRNA sponges<sup>118,119</sup> and one study showed that circZNF609 is associated with ribosomes and therefore could

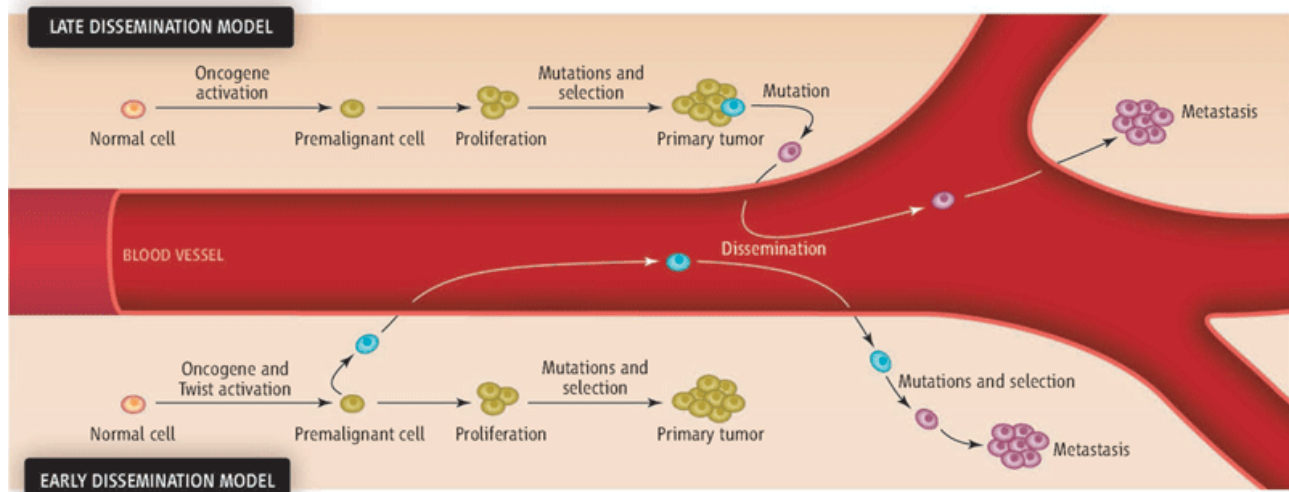
encode for protein<sup>86</sup>. First of all, circZNF609 was reported to be down-regulated in Hirschsprung disease compared with normal bowel tissues. Hirschsprung disease is a developmental defect of the enteric nervous system characterized by lack of enteric neurons in the distal hindgut<sup>120</sup>. This is because the migration of the nerve cells from neural crest is not complete and therefore the activity of colon is not normal. Indeed, this study showed that depletion of circZNF609 inhibited cell proliferation and migration. Mechanistically, circZNF609 may act as a sponge for miR-150-5p to modulate the expression of AKT3<sup>118</sup>. Second, circZNF609 was reported to be upregulated during glucose and hypoxia stress *in vitro* and *in vivo*. In contrast to the previous study, knockdown of circZNF609 increases endothelial cell migration and therefore increase tube formation. To understand its mechanism, it was shown that circZNF609 sequesters and inhibits miR-615-5p activity, leading to increasing expression of MEF2A<sup>119</sup>. Further study showed that circZfp609 (and human circZNF609) is downregulated during muscle differentiation and may encode for a protein. This study proposed that circZfp609 and its protein products may function in regulating myogenesis<sup>86</sup>.

## 1.7 Colorectal cancer progression and metastasis

Colorectal cancer (CRC) accounts for about 1.4 million newly diagnosed cases every year, making it the third most commonly diagnosed malignancy worldwide as reported in 2012<sup>121,122</sup>. Over the past few decades, several studies have revealed several key oncogenes and tumor suppressor genes, especially the APC, KRAS, and p53 genes, are frequently mutated in CRC samples<sup>123,124,125</sup>. For instances, hyperactivation the Wnt pathway, which predominantly occurs through inactivation of the APC gene, drives the transition from normal epithelium to benign adenoma of CRC<sup>126,127</sup>. Liver metastases of CRC occur in about 50% of patients and are the major cause of CRC-related mortality. However, the molecular alterations that support metastases are poorly understood<sup>128,129</sup>.

CRC metastases were first proposed to progress through a clonal tumor evolution process which invasive carcinomas arise from adenomas through the sequential accumulation of mutations in several key genes<sup>130</sup>. Indeed, several mutational profiles of metastases in comparison to primary tumors have shown high concordance between both processes<sup>131,132</sup>. However, conflicting results have been recently reported in some studies showing high discordance mutational rate. For examples, one studies show that in the ten discordant cases reported, 50% of patients whose primary tumor was wild-type had a

mutation in the metastasis while other 50% of patients showed a mutation in primary tumor but not in metastases<sup>133,134</sup>. Indeed, several studies, especially in breast cancer, have recently challenged the generally accepted model for tumor progression through clonal evolution. For examples, there is evidence that metastases can originate parallel with primary tumors<sup>135,136</sup> given that there are several early origins of metastatic lineages already in primary tumors<sup>137,138</sup>.



**Fig. 1.4: Stepwise (late dissemination) and parallel (early dissemination) models of tumor progression during cancer metastasis.** In the stepwise progression, metastases are disseminated at the late stage of tumour development and therefore, the late-disseminated cells and primary tumors are genetically similar. On the other hand, in the parallel progression, the disseminated cells are evolved early and accumulate different genetic and epigenetic alterations and consequently diverge from primary tumors. Understanding the origin of cancer dissemination and metastasis cascades may contribute to a better diagnosis and design of effective therapies. Figure adapted from<sup>139</sup>.

With the development of RNASeq technologies, abnormalities of non-coding RNA expression were found to be associated with CRC pathogenesis. Indeed, accumulating evidence has suggested that miRNAs and lncRNAs are involved CRC metastasis by acting as epigenetic, splicing, transcriptional or post-transcriptional regulators<sup>140,141</sup>. Very recently, several circRNAs were found to be differentially expressed in different stages of CRC suggesting that circRNAs could also play important roles in the progression and metastasis of CRC. Therefore, a better understanding of the expression and function of these non-coding RNA molecules might provide opportunities for understanding the early versus late dissemination mechanisms of CRC metastasis.

## 1.8 Aim of this thesis

Several studies showed the important involvement of lncRNAs and miRNAs during colorectal cancer progression and metastasis. However, at the time of starting this PhD thesis, nothing was known about the roles of circRNAs in development and cancer, especially colorectal cancer. The first aim was to profile the differentially expressed circRNAs in colorectal cancer progression and metastasis. RNAseq was performed using HT29.hCG.Luc intrasplenic SCID mouse model.

With the set of several circRNAs showing differential expression in primary tumors or liver metastases from RNAseq, further investigation revealed that one specific circRNA called circZNF609 is upregulated during cancer progression whereas its linear counterpart, ZNF609 mRNA is not. A further aim of this thesis was to characterize circZNF609 and understand its mechanism in colon cancer progression.

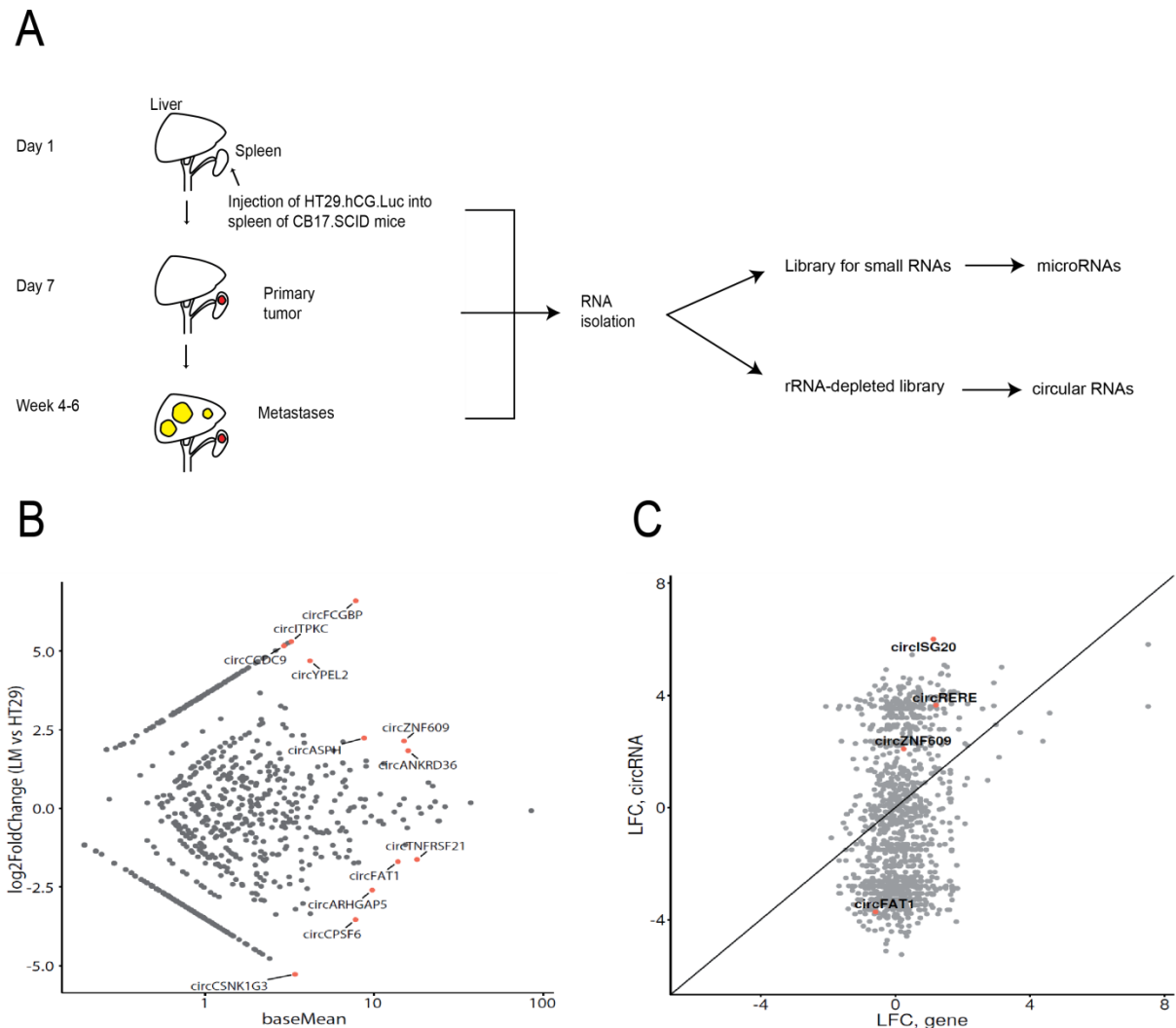
During the course of this PhD thesis, Irene Bozzoni group showed that circZNF609 is associated with ribosomes and may encode for proteins. Therefore, the last aim was to understand how the translation of circZNF609 occurs. This could shed lights on understanding the initiation mechanism of circRNA translation, which is poorly understood compared to the well-studied mRNA translation.

## 2 Results

### 2.1 Mouse model and RNAseq analysis for detection of circRNAs

To identify circRNAs, we generated a mouse model for colorectal cancer development and metastasis (Fig. 2.1 A). Specifically, the human colon cancer cell line HT29 expressing both luciferase and human chorionic gonadotropin (HT29.hCG.Luc) were intrasplenically injected into 6 weeks-aged female CB17 severe combined immunodeficient (CB17.SCID) mice. HT29.hCG.Luc, HCT116.hCG.Luc and CMT93.hCG.Luc were generated as described previously<sup>142</sup>. In this study, we took advantages of bioluminescence (using luciferase) and  $\beta$ -hCG as noninvasive quantification method for monitoring tumour burden. As reported previously in several studies, the relative levels of these surrogate tumour markers were proved to be a stable indicator of tumour burden<sup>143</sup>. Primary tumors and liver metastases were resected about 4-6 weeks after injection when bioluminescence levels reached about 500.000 photons/s. The RNA from HT29.hCG.Luc, primary tumors and liver metastasis was extracted. After quality control of the RNA using Bioanalyzer, miRNA library was prepared using the TrueSeq Small RNA Sample Prep while rRNA-depleted transcriptome library was prepared using the NuGEN paired-end strand-specific RNASeq system.

Bioinformatics analysis for identification and quantification of circRNA expression was performed in collaboration with the laboratory of Nikolaus Rajewsky. We followed methods of identifying the head-to-tail splice junction from RNA sequencing reads as described before<sup>26</sup>. Using this strategy, we have identified several circRNAs differentially expressed in our mouse model (Fig. 2.1 B). Interestingly, several circRNAs showed differential expression while their cognate mRNAs did not change (Fig. 2.1 C). One of the circRNA candidates we would like to characterize further is circZNF609, which is upregulated in primary tumor and liver metastasis from our mouse model.

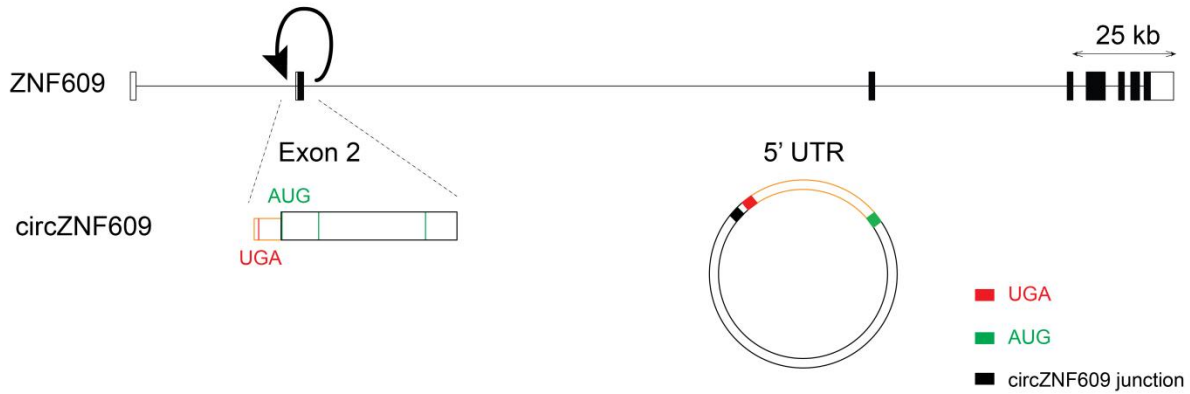


**Fig. 2.1. Identification of circRNAs from HT29.hCG.Luc/SCID mouse model.** A) Mouse model for miRNA and circRNA library preparation. B) Differential expression of circRNAs in liver metastasis versus HT29.hCG.Luc. The total number of reads supporting a particular head-to-tail junction was used as an absolute measure of circRNA abundance. C) Scatter plot of circRNA expression and linear transcript expression in liver metastasis versus HT29.hCG.Luc. Expression of some circRNAs is independent from its host gene.

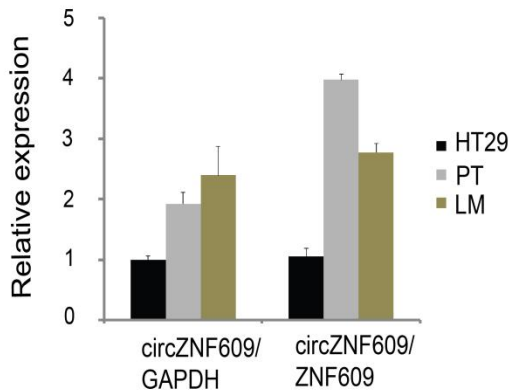
## 2.2 Characterization of circZNF609

CircZNF609 is a 874 nucleotides (nts) long RNA and represents a circularized exon of ZNF609 mRNA. It is generated by a back-splicing event of exon 2 of ZNF609 (Fig. 2.2 A). CircZNF609 shares AUG with its cognate mRNA ZNF609. Interestingly, we found a stop codon upstream of the circZNF609 AUG which could generate a complete open reading frame (ORF) upon circularization (Fig. 2.2 A).

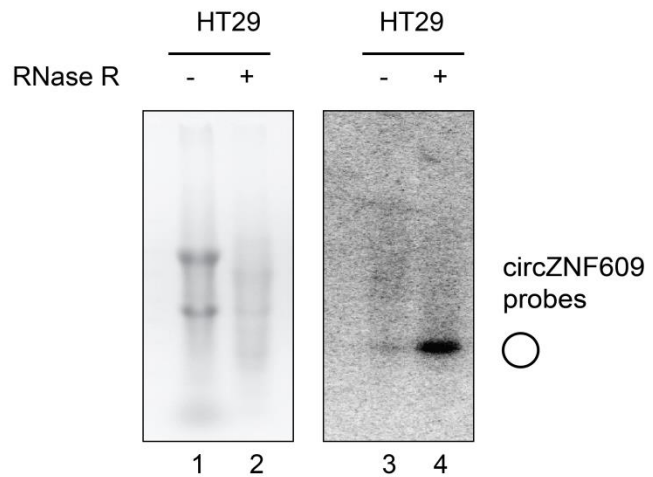
A



B



C



**Fig. 2.2. Characterization of circZNF609.** A) Scheme indicates the ZNF609 locus with circZNF609 generated from its exon 2. The circularized exon 2 of circZNF609 is located between very long introns as observed for several circRNAs. The stop codon (red color) locates in front of the start codon AUG (green color). Upon circularization, circZNF609 forms a complete ORF and a 5' UTR (yellow color) locates between the stop and start codon. B) Relative expression of circZNF609 in primary tumor (PT) and liver metastasis (LM) validated by qPCR. The relative expression of circZNF609 was normalized to the expression of GAPDH and its host gene, ZNF609 mRNA, respectively. C) Total RNA from HT29.hCG.Luc treated with RNase R (lane 2) or mock treated (lane 1) were loaded on agarose gel. Northern blot analysis of circZNF609 after RNase R treatment (lane 4) showed significant levels of circZNF609.

Validation of the expression of circZNF609 in our mouse model by qPCR confirmed that circZNF609 is upregulated during cancer progression while the expression of ZNF609 mRNA is unchanged (Fig. 2.2B). To further confirm the circularity of circZNF609, we used



RNase R, a 3' to 5' exoribonuclease<sup>35</sup>, to digest all linear RNAs. Northern blot showed that circZNF609 is resistant to RNase R treatment confirming the circular nature of this molecule (Fig. 2.2 C). Further characterization revealed that circZNF609 is not enriched in poly (A) (+) fraction as observed for the ZNF609 mRNA (Supplemental Fig. 5.9 A), indicating that circZNF609 is indeed a circle.

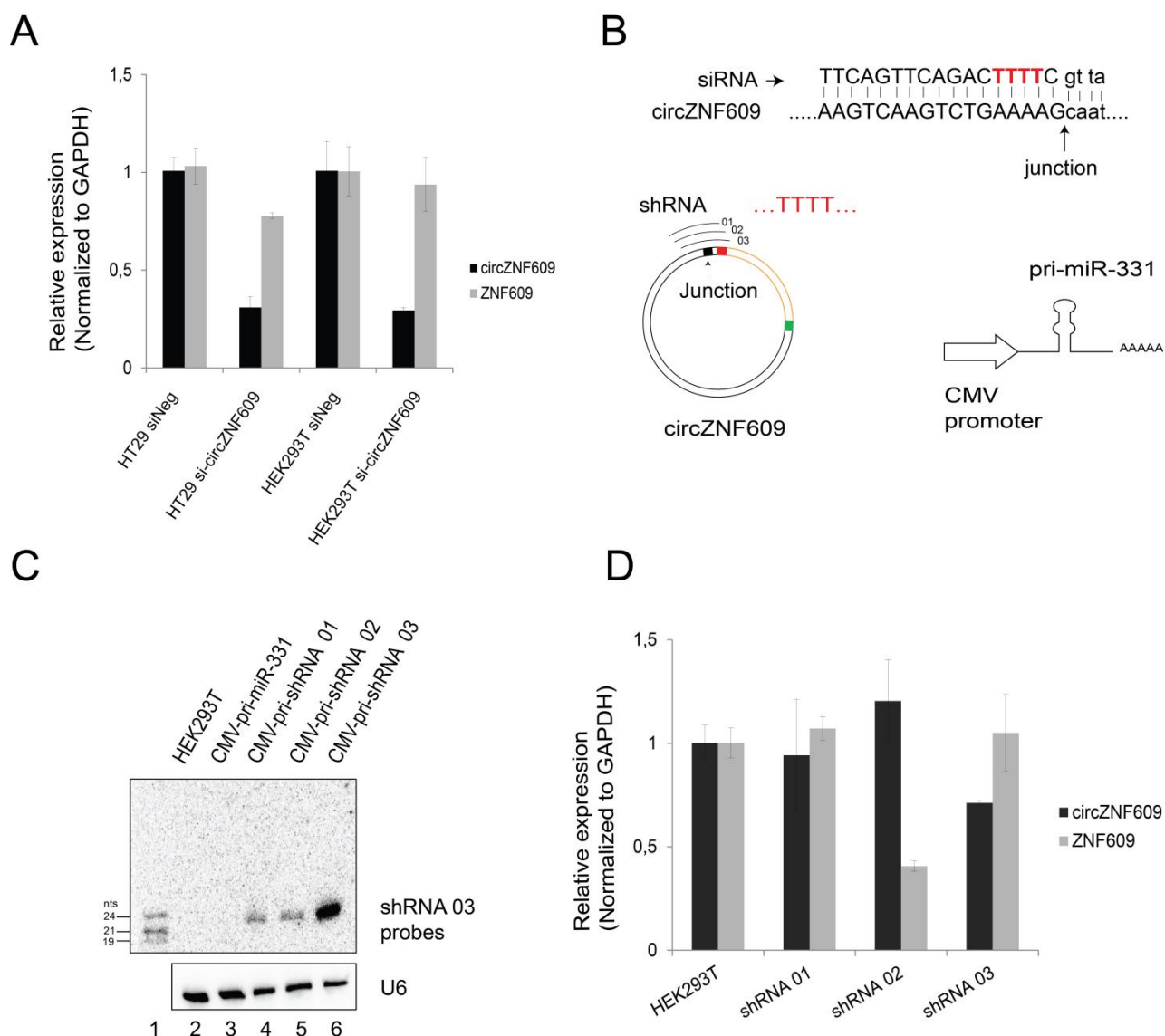
Taken together, we have identified and confirmed that circZNF609 is a circular RNA and is upregulated during colorectal cancer progression and metastasis. Further functional experiments would be needed to understand the significance of circZNF609 upregulation in CRC progression.

## 2.3 Generation of shRNA constructs targeting circZNF609

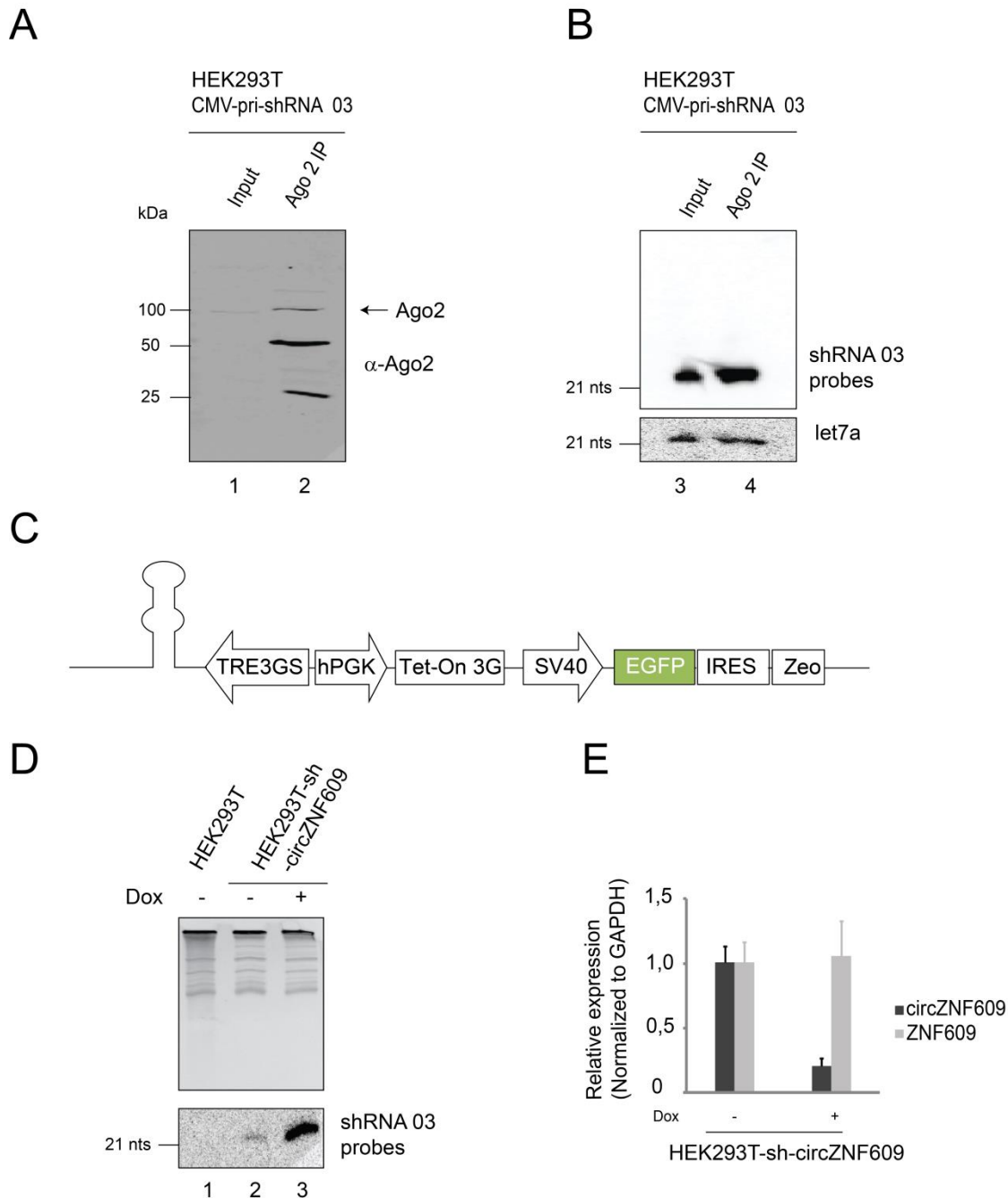
To investigate the potential function of circZNF609 on CRC development, we performed the knockdown experiments. Therefore, we designed an siRNA for knocking down of circZNF609. Since circZNF609 shares its sequence with ZNF609 mRNA, the siRNA was designed to target the junction of circZNF609 (Fig. 2.3 B). qPCR from HT29.hCG.Luc and HEK293T cell lines transfected with the designed siRNA showed that this siRNA can specifically target circZNF609 without affecting the expression of ZNF609 (Fig. 2.3 A).

To study the impact of circZNF609 on primary tumor and metastasis development *in vivo*, we wanted to generate a stable HT29.hCG.Luc cell line, which itself can produce siRNAs to knock down circZNF609 upon doxycycline induction. This is critical because we can manipulate the expression of circZNF609 after the cells have been implanted into the mice.

To achieve silencing of circZNF609 via shRNA, a pSuper vector was used. Correct shRNA synthesis and processing was verified by northern blot using total RNA isolated from HEK293T cells. At first, it is worth mentioning that the stop signal for transcription of the H1 promoter, which is RNA polymerase III (Pol-III) promoter, is determined by five Ts in a row (TTTTT). Unfortunately, due to the fact that the designed sequences had to comprise a number of nucleotides in the junction of circZNF609, all of our constructs contained four Ts in a row in front of the stop signal likely causing premature abortion of Pol-III transcription (Fig. 2.3 B). As expected, northern blot analysis revealed no significant expression of the siRNA from the pSuper constructs (data not shown).



**Fig. 2.3. Generation of shRNA constructs targeting circZNF609.** A) HEK293T or HT29.hCG.Luc was transfected with siRNA targeting circZNF609. 3 days post-transfection, the RNA was harvested for qPCR analysis. B) siRNA strategy for targeting circZNF609. Alignment of the siRNA with junction of circZNF609 is indicated. All 3 designed sequences spanning the junction of circZNF609 contain four Ts in a row in front of the stop signal, therefore lead to premature abortion of Pol-III transcription. Alternatively, artificial microRNA (amiRNA) construct based on pri-miR-331 driven by Pol-II promoter was developed. C) HEK293T was transfected with the constructs consecutively expressing pri-miR-331 or different amiRNAs. 2 days post-transfection, RNA was harvested for northern blot analysis using probes targeting the desired sequence of shRNA 03. D) HEK293T was transfected with these amiRNAs constructs using Lipofectamine 2000. 3 days post-transfection, the RNA was harvested for qPCR analysis. The relative expression of circZNF609 and ZNF609 was normalized to the expression of GAPDH



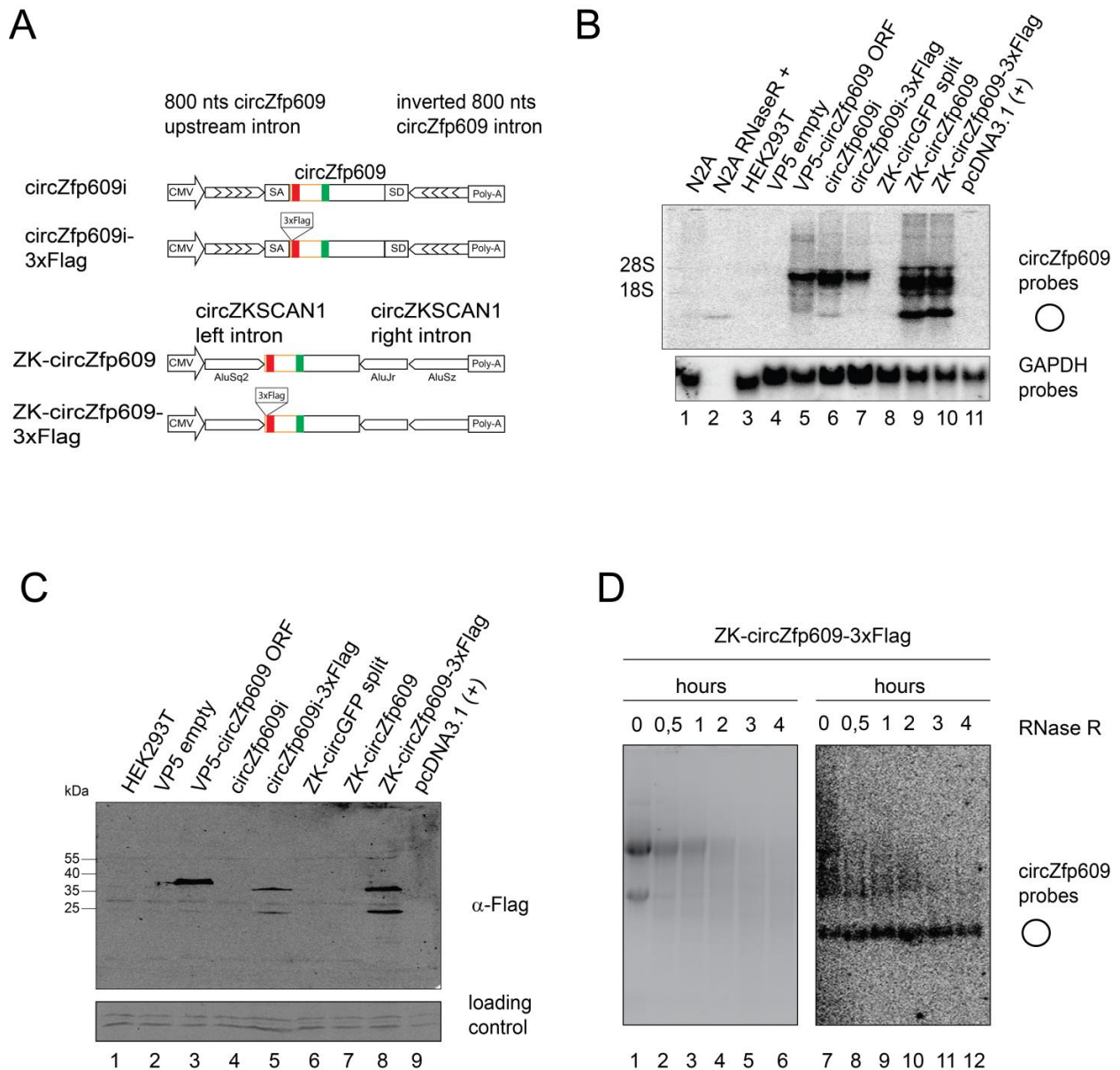
**Fig. 2.4. Characterization of the amiRNA 03 for establishing the inducible cell line.** A) HEK293T was transfected with amiRNA 03. 2 days post-transfection, the cells were harvested for Ago2 IP using the anti-Ago2 antibodies (clone 11A9). B) The RNA from (A) was run on 12% Urea PAGE for northern blot analysis. Let7a was used as positive control. C) Inducible vector for expression of shRNA targeting circZNF609. D) HEK293T-sh-circZNF609 was induced with doxycycline for 3 days. Afterwards, the cells were harvested for northern blot analysis to confirm the expression of amiRNA 03. E) Relative expression of circZNF609 and ZNF609 in HEK293T-sh-circZNF609 upon doxycycline induction validated by qPCR. The relative expression of circZNF609 and ZNF609 was normalized to the expression of GAPDH. Dox: Doxycycline

To overcome this problem, we designed an artificial microRNA (amiRNAs) construct based on pri-miR-331 that is driven by the CMV promoter, which is RNA Pol-II promoter. We designed 3 shRNAs spanning the junction of circZNF609 (Fig. 2.3 B). Analysis by northern blot revealed that all three antisense strands were produced in a correct manner and showed good expression of around 21 nts (Fig. 2.3 C). The data also showed that the probe shRNA 03 binds not only amiRNA 03 (Fig. 2.3 C, lane 6) but also the other two amiRNAs (Fig. 2.3 C, lane 4, 5). This effect is based on the high sequence similarities of the constructs due to the low variety in the design of the different shRNAs, as they were designed in a manner to overlay the junction of the circZNF609. qPCR analysis of circZNF609 expression three days post transfection of these amiRNA constructs in HEK293T cells showed that amiRNA 03 leads to downregulation of circZNF609 expression without affecting the expression of ZNF609 (Fig. 2.3 D). Of note, by moving few nts from the junction, the amiRNA 02 targets only ZNF609, but not circZNF609 (Fig. 2.3 D). Therefore, amiRNA 03 was selected for further analysis. Further investigation revealed that the generated siRNAs from amiRNA 03 are loaded into Ago2 (Fig. 2.4 A, B). This implied that the siRNAs generated from this pri-miR-331 are functional.

To generate the stable inducible cell lines expressing shRNA constructs, the CMV promoter (Fig. 2.3 A) was replaced by the inducible TRE3GS promoter (Fig. 2.4 C). Using this vector, we could obtain stable cell lines HEK293T (Fig. 2.4 D, E) expressing shRNA (HEK293T-sh-circZNF609). Upon induction with doxycycline, HEK293T-sh-circZNF609 expresses siRNA targeting circZNF609 (Fig. 2.4 D). Consequently, upon doxycycline induction we could knock down circZNF609 without affecting the expression of ZNF609 (Fig. 2.4 D).

Collectively, we could develop a construct for specifically targeting circZNF609 without targeting ZNF609 in an inducible manner. Then, we could generate an inducible stable cell line HEK293T-sh-circZNF609 for knocking down of circZNF609 upon doxycycline induction. This is fundamental for establishing inducible stable cancer cell lines to study circZNF609's roles during colorectal cancer progression.

## 2.4 Methods for overexpression of circZNF609



**Fig. 2.5. Generation of overexpression constructs for circZfp609.** A) Scheme of constructs generated for overexpression of circZfp609, the mouse ortholog of human circZNF609. Specifically, circZfp609 was produced from plasmids circZfp609i as described in Hansen et al., 2013 (Kjems lab) or plasmid ZK-circZfp609 using circZKSCAN1 intron as described in Kramer et al., 2015 (Willusz lab)<sup>52</sup>. 3xFlag was inserted in front of the stop codon of circZfp609 to generate circZfp609i-3xFlag and ZK-circZfp609-3xFlag plasmids. B) HEK293T cells transfected with different overexpression constructs was analysed by northern blot. C) Western blot of HEK293T cells transfected with different constructs from (B) detected by anti-Flag antibodies. D) The RNA from HEK293T cells transfected with ZK-circZfp609-3xFlag plasmids were harvested and subjected to RNase R digestion with different time points.

To be able to overexpress circRNAs and to study phenotypes resulting from overexpressed circZNF609, we used different methods based on published data<sup>70,52</sup>. First, we decided to overexpress the mouse version of the human circZNF609 (circZfp609). This is because we wanted to check the fidelity of overexpressed murine circZfp609 in human HEK293T cell lines. Northern blot analysis confirmed expression of the overexpressed circZfp609 when comparing with the endogenous circZfp609 (Fig 2.5 B). Moreover, we also confirmed the conclusion from the Willusz lab that the method developed by Willusz lab<sup>52</sup> is much more efficient in producing circRNAs compared to the method from Kjems lab<sup>70</sup>.

To further confirm the fidelity of overexpressed circRNAs, we digested the RNA with RNase R at different time points and observed that the circZfp609 was stable while other linear RNAs and processing intermediates were digested (Fig. 2.5 D). To have a solid conclusion about the correct circRNA expression, we validated the overexpressed murine circZfp609 in HEK293T cells by Sanger sequencing. Even though circZNF609 is highly conserved, there are few mismatches in the sequence spanning the junction of circZfp609 and circZNF609 (Supplemental Fig. 5.9 C). Therefore, Sanger sequencing confirmed the expression of murine circZfp609 in human HEK293T cells (data not shown).

Taken together, we have successfully established molecular tools to overexpress and deplete circZNF609. These tools are necessary to understand the mechanisms of circZNF609.

## **2.5 CircZNF609 potentially encodes for proteins**

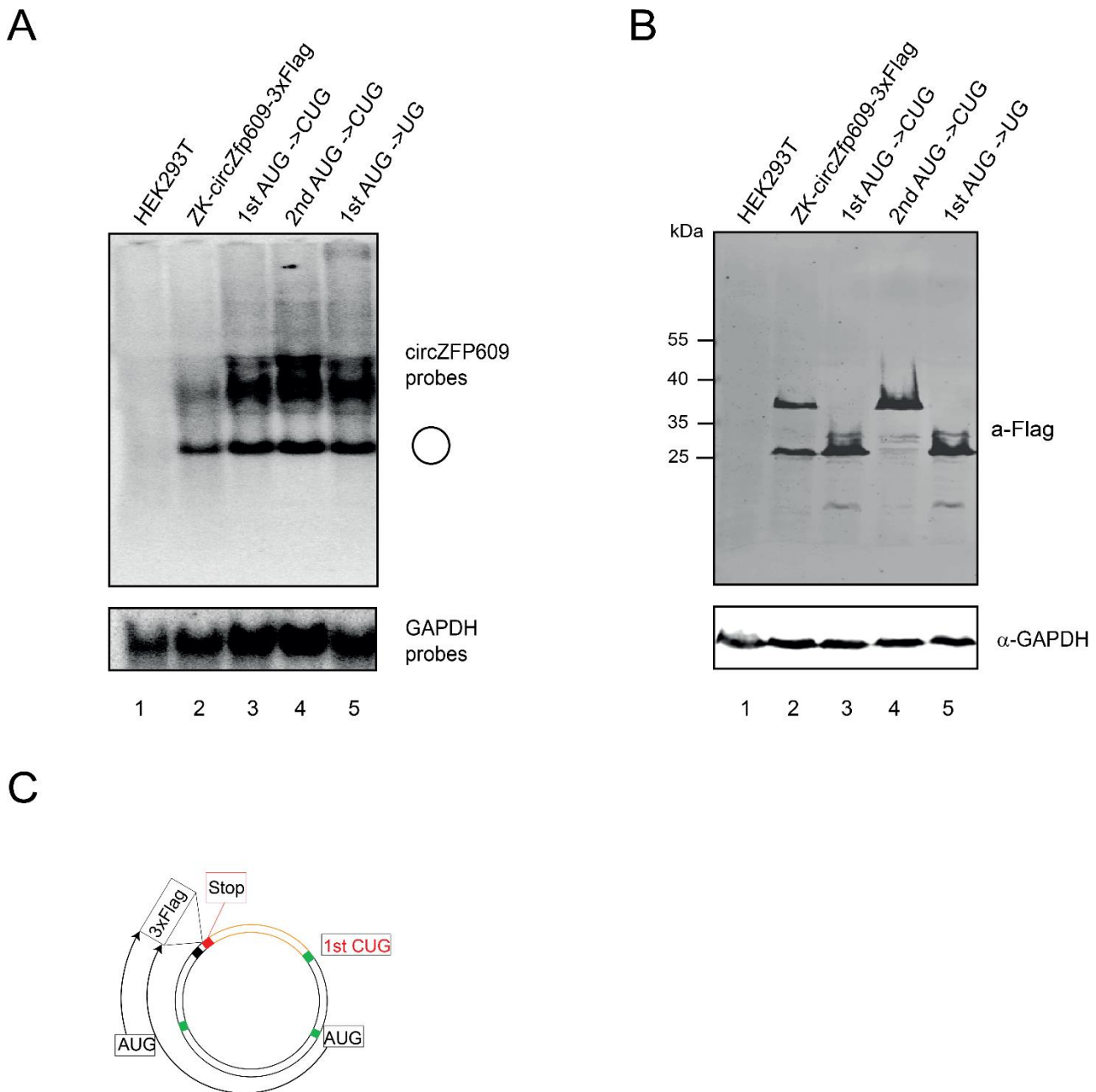
### **2.5.1 Flag reporter constructs and mutagenesis**

During the course of this study, Irene Bozzoni's lab reported that circZNF609 is bound to ribosomes and therefore potentially encodes for small protein(s)<sup>86</sup>. To investigate this hypothesis, a 3xFlag tag was inserted in front of the stop codon of circZfp609. Northern blot confirmed that circZfp609 with 3xFlag tag is generated (Fig. 2.5 B, lane 10). This construct was then used and validated further in the mentioned publication by the Bozzoni lab<sup>86</sup>. Consistent with the data from the Bozzoni lab, western blot with anti-Flag antibodies showed two protein products implying that two AUGs could be used from circZfp609 (Fig. 2.5 C). Indeed, two AUGs are present that would explain the two protein bands (Fig. 2.5 C). Interestingly, we used two different vectors containing different intron sequences

(upstream and downstream of circZfp609) for overexpression of circZfp609-3xFlag (Fig 2.5 A). Both overexpression constructs gave us two protein products confirming that both protein products are generated from circZfp609-3xFlag. Notably, the relative expression of circZfp609-3xFlag transcripts was less produced from circZfp609i-3xFlag plasmids (Fig 2.5 B, lane 7) than from ZK-circZfp609-3xFlag plasmids (Fig 2.5 B, lane 10). Consequently, the two protein products were less produced from circZfp609i-3xFlag plasmids (Fig 2.5 C, lane 5) compared to ZK-circZfp609-3xFlag plasmids (Fig 2.5 C, lane 8).

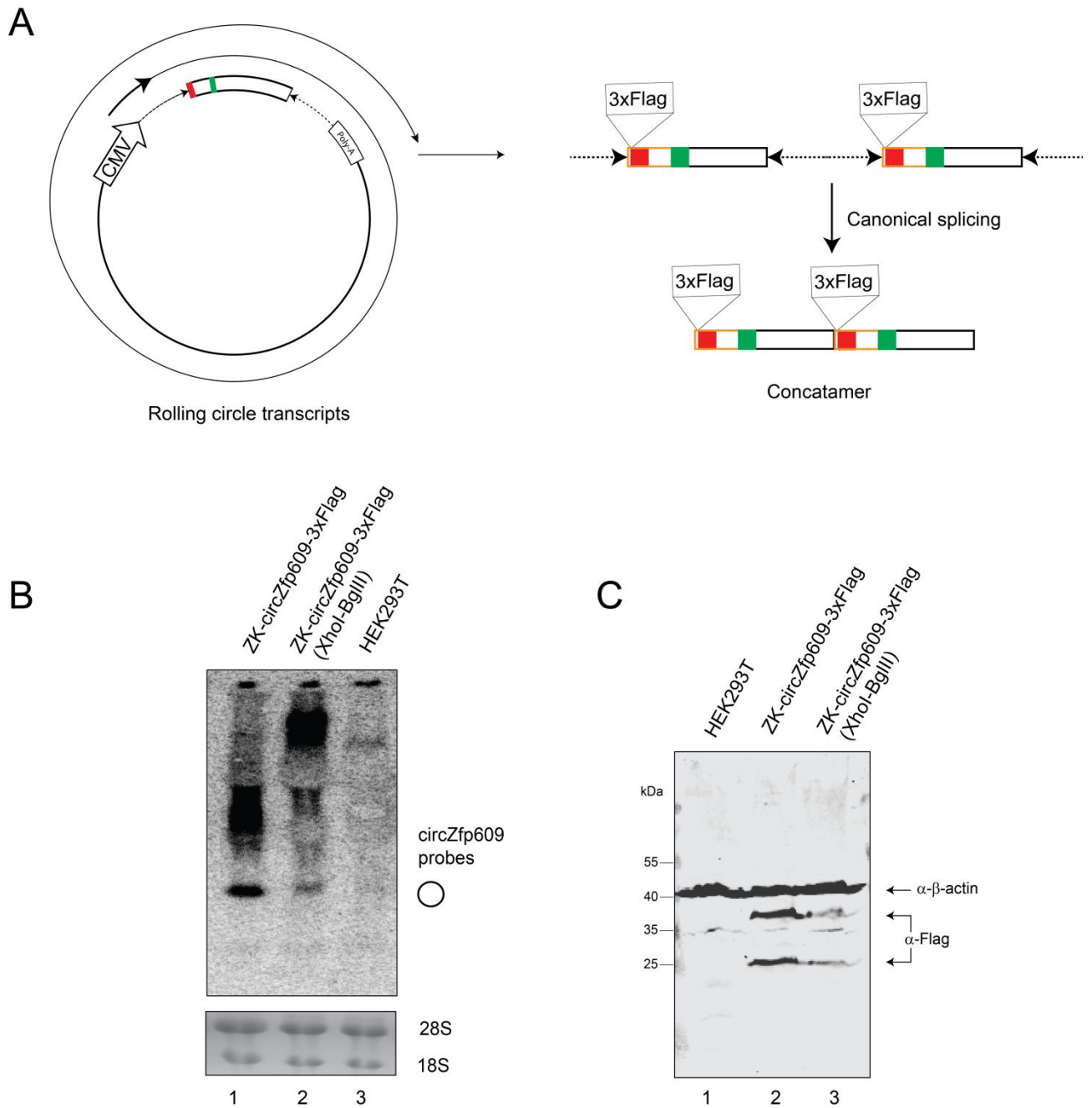
To prove that two AUGs from circZfp609 were used for translation, the first or second AUG was mutated to CUG (Fig 2.6 A, lane 4, 5) or the first AUG was mutated to UG (Fig 2.6 A, lane 6). Northern blot confirmed the expression of these mutants (Fig 2.6 A) and western blot confirmed the usage of those AUG accordingly (Fig 2.6 B). Specifically, when the first AUG was mutated to CUG (Fig 2.6 B, lane 3) or mutated to UG (Fig 2.6 B, lane 5) we did not observe the expression of upper protein products. When the second AUG was mutated to CUG, the lower protein products were not produced (Fig 2.6 B, lane 4). Interestingly, we observed that a third AUG is used when the first AUG is mutated (Fig 2.6 B, lane 3, 5 and Fig 2.6 C). While two AUGs are normally used from circZfp609, all three AUGs are used from circZfp609 when we lengthened the sequence of 5'UTR of circZfp609 (Fig. 2.10 A and Fig. 2.10 B, lane 3). In this case, we added 49 nts unrelated sequence (49 nts of  $\beta$ -globin UTR) just in front of the first AUG. Western blot showed three protein bands corresponding to three observed AUGs on circZfp609 (Fig. 2.10 B, lane 3). Further experiments are required to understand the mechanism of how three AUGs are used for translation.

It was reported that, in some plasmids, rolling circle transcription can occur due to the fact that the RNA polymerase can sometimes bypass the termination signal. This leads to production of several unwanted transcripts, including a precursor transcript containing tandem repeats of the circularized exon. These transcripts are suggested to be spliced canonically and a linear concatamer of the circularized exon are generated<sup>144</sup>. In this case, the concatamer represents the junction of circRNA, which could be used for translation (Fig. 2.7 A) and therefore the protein products that we observed could be unspecific.

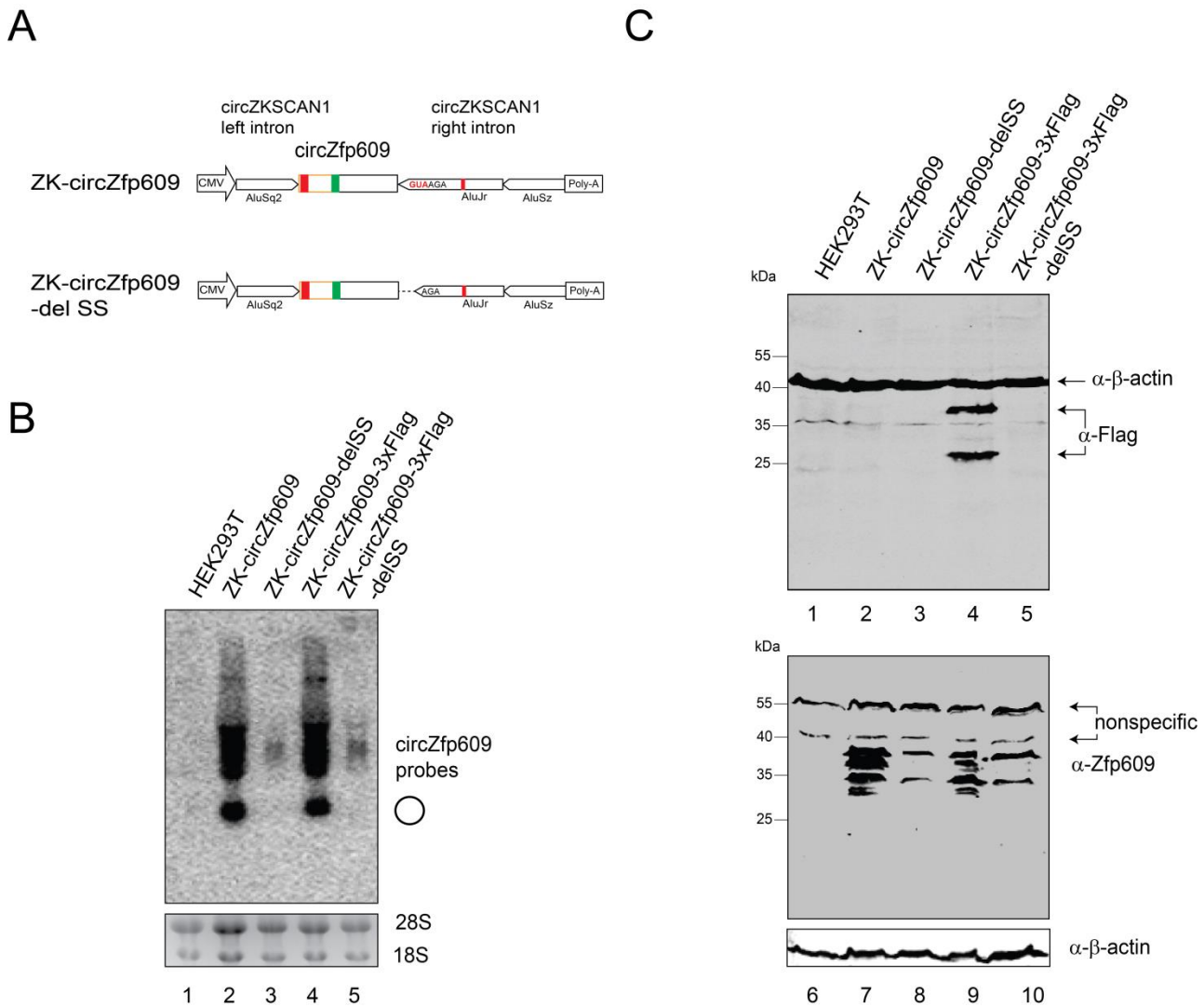


**Fig. 2.6. Mutagenesis of AUGs for understanding translation of circZfp609.** A) Northern blot of RNA isolated from HEK293T cells transfected with different overexpressed circZfp609 AUG mutants. GAPDH served as loading control. B) Western blot of the mutants from (A) detected by anti-Flag antibodies. The third AUG is used only when the first AUG is mutated to ACG (lane 3) or to UG (lane 5). C) Analysis of circZfp609 sequence predicts 3 AUGs in-frame with 3xFlag and thus could be detected by anti-Flag antibodies.





**Fig. 2.7. Confirmation of circZNF609 translation.** A) Scheme of proposed rolling circle transcription and generation of linear concatamer. B) Northern blot of transfected fragments from plasmids ZK-circZfp609-3xFlag digested with XhoI and BglII. C) Western blot of the HEK293T cells transfected with according digested ZK-circZfp609-3xFlag.



**Fig. 2.8. Deletion of splice site abrogates circZfp609 production and its corresponding proteins.** A) Scheme represents deletion of splice sites (SS) for disruption of circZfp609 production. Shown here are examples of ZK-circZfp609. Same delSS was performed with ZK-circZfp609-3xFlag. Specifically, in this delSS mutant, 3 nts SS (GUA) in the intron of circZKSCAN1 was deleted. B) Northern blot of all transfected mutants in HEK293T cells. C) Western blot of HEK293T transfected with SS-depleted mutants by anti-Flag and anti-Zfp609 antibodies.

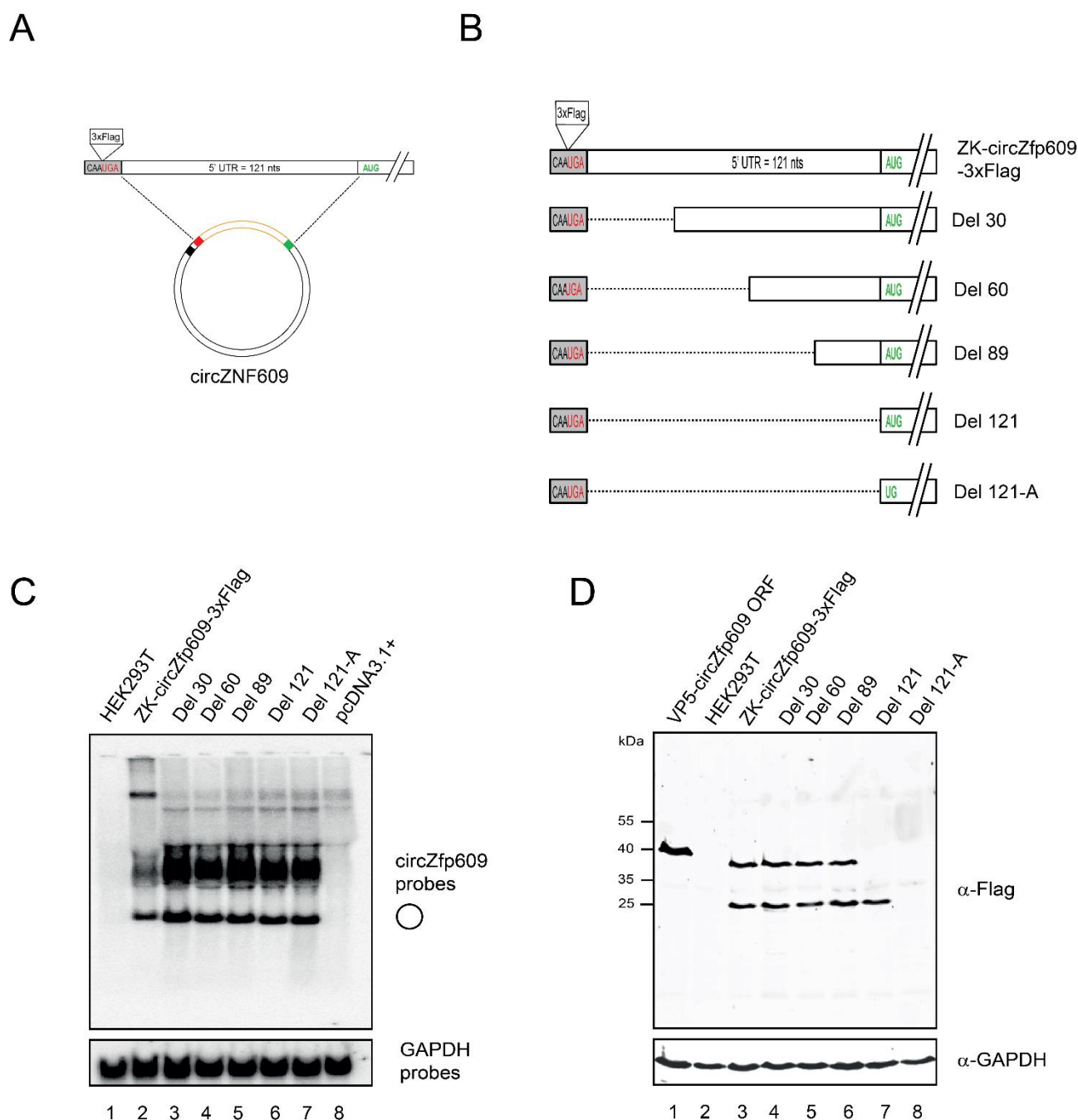
Even though the pcDNA3.1 (+) plasmids contain a strong bGH poly (A) stop signal, we wanted to test if the proteins generated from this plasmids are from the concatamer. To address this, the plasmids were digested into half by BglII and XhoI. The fragments containing the circZfp609 flanking by circZKSCAN1 intron were purified from Agarose gel. Northern blot showed the expression of circZfp609 from these fragments (Fig. 2.7 B). Western blot from the transfected fragments gave 2 protein products recognized by the

anti-Flag antibodies (Fig. 2.7 C). This confirmed that the protein products are generated from the circRNAs, not from the concatamer.

To further confirm that the protein products were produced from circRNAs, we disrupted the production the circRNAs by deleting 3 nts of the splice site (GUA) from the circZKSCAN1 right intron (Fig. 2.8 A). Since GU and AG dinucleotides are important splice sites for splicing<sup>145</sup>, deletion of 3 nts GUA may abrogate the back-splicing of circZfp609. Indeed, northern blot showed that circRNAs were not made from the cells, only the linear RNAs (Fig. 2.8 B). Consequently, we did not observe any proteins from the ZK-circZfp609-3xFlag-delSS mutant using anti-Flag antibodies (Fig. 2.8C) confirming that the proteins are indeed produced from circRNAs, not from any linear RNA, which is generated during the circRNA biogenesis process.

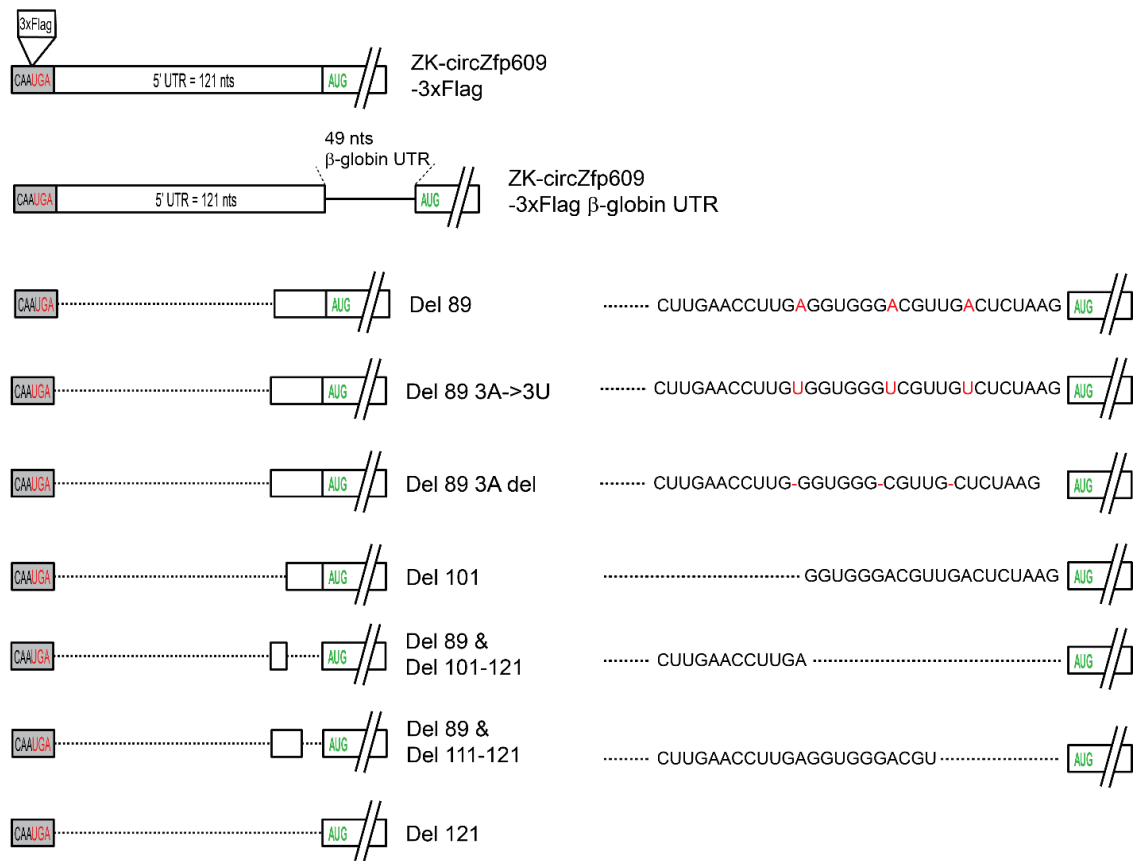
To search for a potential IRES sequence which may contribute to the recruitment of ribosomes for translation, we analyzed the 5'UTR of circZfp609 since this is the region where IRES elements are often located in other RNAs (Fig. 2.9 A). Mutagenesis was performed with different truncations of the 5'UTR (Fig. 2.9 B). Importantly, deletion of 121 nts 5'UTR (Del 121) showed only one protein band while Del 121-A did not give us any protein products recognized by the anti-Flag antibodies (Fig. 2.9 D). Of note, northern blot confirmed that all truncated mutants produce circZfp609-3xFlag in HEK293T cells (Fig. 2.9 C).

Deletion of 89 nts 5'UTR (Del 89) still gave us 2 protein bands (Fig. 2.9 C) implying that the IRES sequence may locate in the last 32 nts of 5'UTR. Several other mutants of these 32 nts of 5'UTR were generated (Fig. 2.10 and Supplemental Fig. 5.11 and 5.12) and western blot revealed different protein products for different mutants (Fig. 2.10 B). Specifically, the Del 101 mutant still gave two protein products while the other mutant (Del 89 & Del 111-121, Fig. 2.10 B, lane 8) did not produce these two protein products, but a fainter higher molecular weight protein product. The result from these mutants implied that 32 nts of 5'UTR does not contain the IRES element and consequently, the 5'UTR does not function as IRES for translation of circZfp609. Therefore, the IRES sequence of circZfp609 may locate in its coding sequence. Further experiments and mutagenesis would be required to conclude this hypothesis.

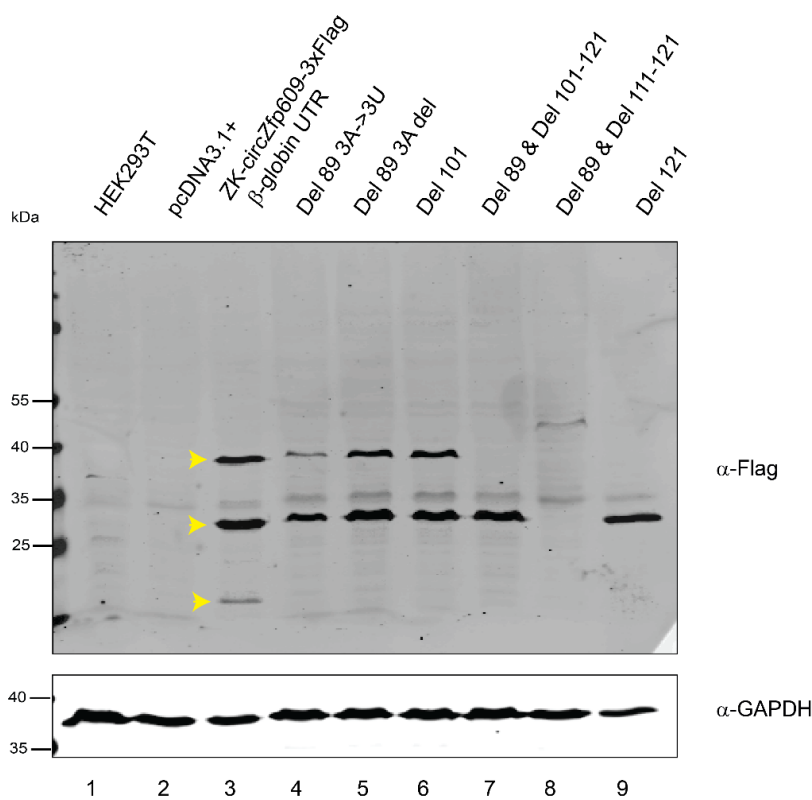


**Fig. 2.9. Mutagenesis of Zfp609 UTR for studying its translation mechanism.** A) Scheme of 121 nts of circZfp609 5' UTR in between the stop codon UGA and the first start codon AUG. B) Truncated mutants of circZfp609 5' UTR. C) Northern blot of HEK293T cells transfected with all 5' UTR deletion mutants show expression of overexpressed circZfp609 5' UTR mutants. D) Western blot of samples from C) using anti-Flag antibodies.

A



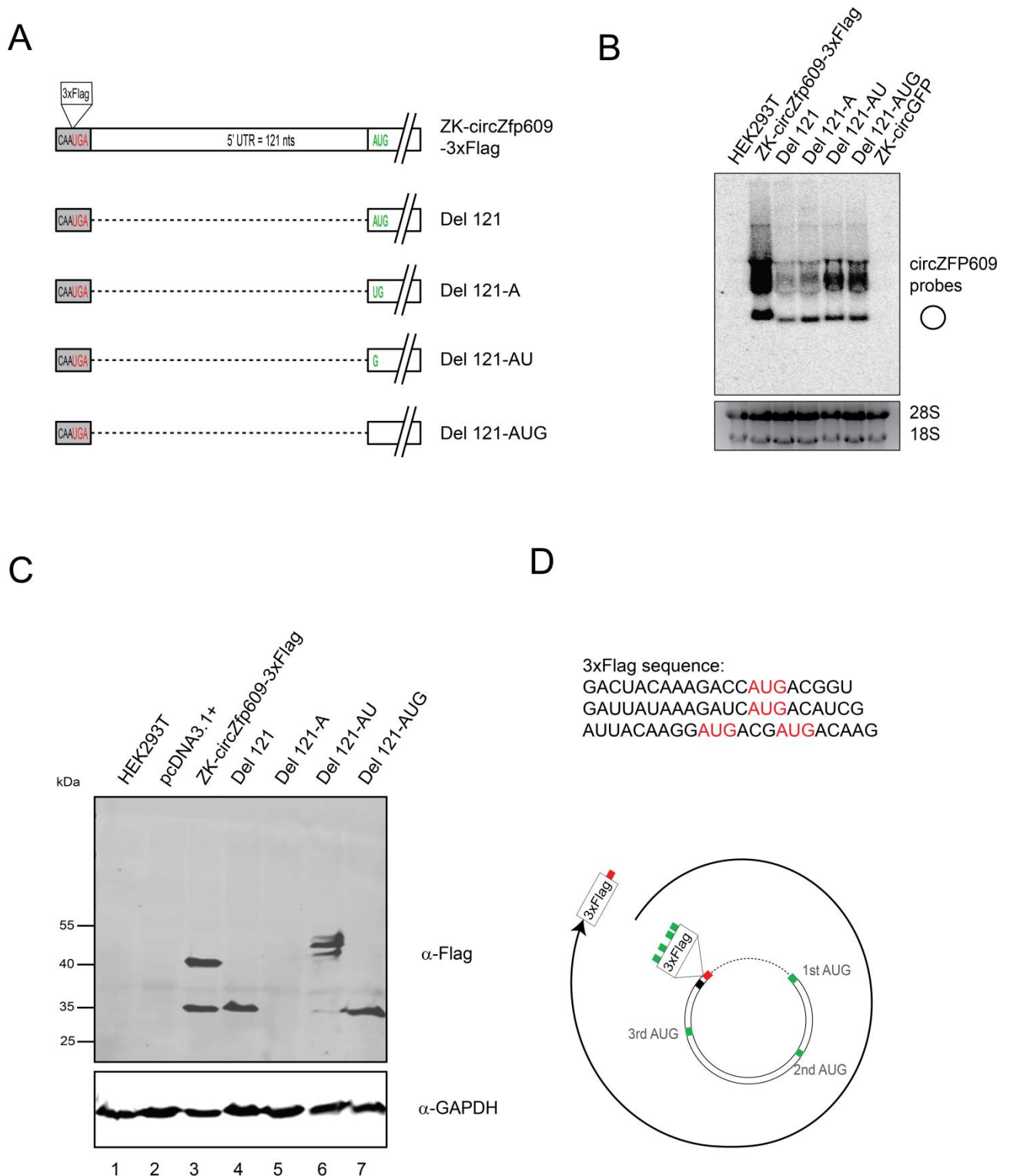
B



**Fig. 2.10. Further analysis of Zfp609 5' UTR.** A) Truncated mutants of circZfp609 5' UTR were generated to understand how circZfp609 can recruit the ribosome. 49 nts of the  $\beta$ -globin UTR was inserted before the AUG of circZfp609. In this case, we observed all 3 AUGs are used for translation. Del 89 represents several A could potentially m6A methylated as suggested from RMbase<sup>146</sup>. All 3 As were mutated to Us or all 3As were deleted. For translation of circRNA, it is very important to delete 3 nts or sequentially each nucleotide because of the frameshift as discussed. B) Western blot of the HEK293T cells transfected with different mutants by anti-Flag antibodies. Arrow indicates the proteins products.

To understand the protein products generated from these above mutants, we proposed that, on a circle, translation of circZfp609 is shifted to different ORFs when the 5'UTR or different parts of the 5'UTR are deleted. To check this hypothesis, we generated other mutants where we deleted sequentially one, two and three nts of Del121 constructs (Fig. 2.11 A). If our hypothesis is correct, we will observe the same protein products from Del 121 and Del 121-AUG mutants. Indeed, northern blot confirmed the expression of all deletion mutants (Fig. 2.11 B) and western blot using anti-Flag antibodies revealed that Del121-AUG give us the same protein bands (Fig. 2.11 C, lane 4) as observed for Del121 (Fig. 2.11 C, lane 7).

Unexpectedly, we observed higher molecular weight protein products of the Del121-AU mutants than the two normal protein products produced from ZK-circZfp609-3xFlag (Fig. 2.11 C, lane 6). The result from this Del121-AU mutant implied that there are other upstream AUGs, which could be used for translation. Importantly, these protein products were recognized by anti-Flag antibodies meaning that the Del121-AU starts translation at upstream AUGs and ends at the 3xFlag (Fig. 2.11 D). Moreover, in this case, we did not observe the two normal protein products as observed for ZK-circZfp609-3xFlag. This could be explained as following: when the upstream AUGs are used for translation, the downstream AUGs (in this case, the two normal AUGs) from circZfp609 are not used any more. Moreover, we did not observe any protein products from Del 121-A mutants (Fig. 2.11 D). In this case, this could be explained by two possibilities: either the upstream AUGs or the two normal AUGs were used for translation but, in both cases, their ORFs did not end at the 3xFlag frame. Therefore, we did not observe any protein products recognized by anti-Flag antibodies. In summary, these mutants confirmed that there was a frame change on circZfp609 when one nucleotide is deleted from Del 121.



**Fig. 2.11. Translation of circZfp609.** A) Mutagenesis of each one nucleotide from Del 121 for understanding its translation. B) Northern blot of HEK293T cells transfected with all truncated mutants showed significant expression of the overexpressed circZfp609. C) Western blot of deletion mutants using anti-Flag antibodies. D) Sequence of 3x Flag contains 4 AUGs, which can be used for translation of circZfp609.

To further understand the protein products generated from Del 121-AU mutant, we analyzed the upstream sequence of its 5' UTR. Unexpectedly, analysis of the sequence from circZfp609-3xFlag revealed four out-of-Flag-frame AUGs in the 3xFlag sequence and we proposed that these four AUGs are used for translation when the normal circZfp609 ORF is shifted (Fig. 2.11 D). Further mutagenesis of these mutants is required to have a solid conclusion about this observation.

Taken together, mutagenesis of 5' UTR of circZfp609 showed that this 5' UTR does not contain IRES elements. Furthermore, the translation of circZfp609 is possible and several AUGs can be used for translation on circZfp609. The mechanisms of how different AUGs were used are still not known and further analysis is required for a final conclusion. All western blots in these experiments were performed using anti-Flag antibodies, further western blot analysis using anti-Zfp609 antibodies is necessary for better understanding of circZfp609 proteins and translation mechanisms.

### 2.5.2 Detection of circZfp609 proteins by anti-Zfp609 antibodies

Since the overexpression of circZfp609 proteins from the mentioned vectors gave several products, it is essential to analyze circZfp609 translation under physiological condition. As a first approach, we analyzed the expression of circZfp609 proteins in a stable cell line using anti-Zfp609 antibodies. For *in vivo* xenograft study (see 2.8), we developed an inducible stable cell line for overexpression of circZfp609 upon doxycycline induction (CMT93-circZfp609) (Fig. 2.12 B). Using anti-Zfp609 antibodies to detect circZfp609 proteins, western blot from this CMT93-circZfp609 stable cell line upon doxycycline induction showed at least 2 protein bands recognized by anti-Zfp609 antibodies (Fig. 2.12 B, lane 4). One protein band has similar size with proposed endogenous protein (Fig. 2.12 B, lane 3). Unexpectedly, we observed two more protein bands which are only expressed in the CMT93-circZfp609 cells upon doxycycline induction (Fig. 2.12 B, lane 3).

To understand the expression of these additional protein bands, we noticed that there are at least two RNA products generated from circZfp609 overexpression vectors (Fig. 2.12 A). Specifically, after transcription, in addition to the circZfp609 transcript, a linear transcript was also produced (Fig. 2.12 A). The linear transcript shares its sequence with circZfp609 transcript. The only difference between circZfp609 transcript and the linear transcript is the usage of stop codon. As mentioned before, circZfp609 transcript uses an upstream stop codon upon circularization. On the other hand, the linear transcript uses a



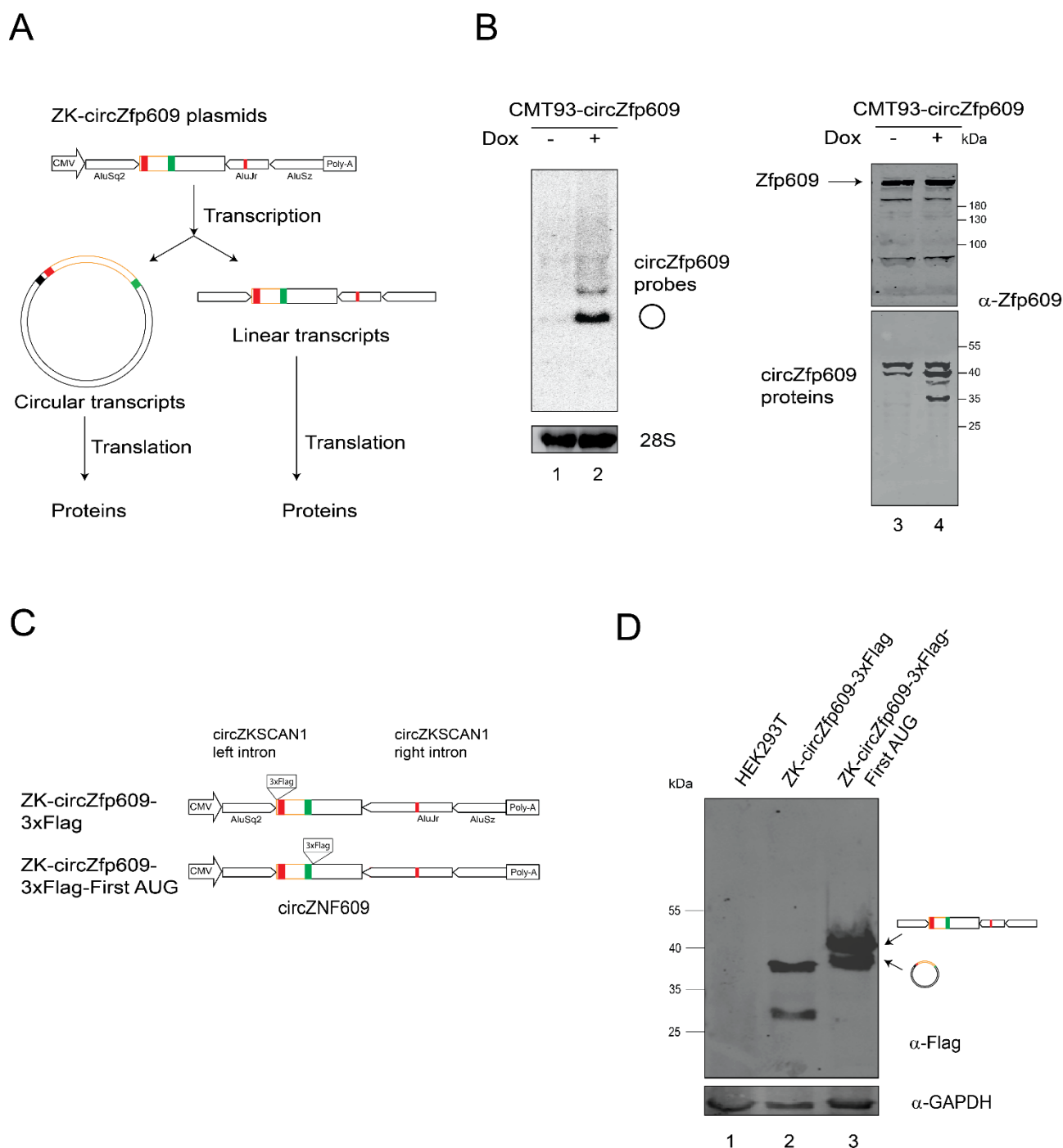
stop codon from the circZKSCAN1 intron, which is downstream of the circZfp609 exon (Fig. 2.12 A). Since both transcripts contain similar sequence, their protein products are both recognized by the anti-Zfp609 antibodies. Therefore, there are more protein products are detected in western blot using the anti-Zfp609 antibodies.

To further prove that the linear transcripts from overexpression vectors could be used for translation, 3xFlag sequence was inserted just after the first AUG of circZfp609 (Fig. 2.12 C) instead of in front of stop codon as described in ZK-circZfp609-3xFlag (Fig. 2.5 A). Western blot using anti-Flag antibodies showed two proteins products corresponding to two transcripts: one belongs to the circRNA transcript and another one belongs to the linear RNA transcript (Fig. 2.12 D).

In the same line with this observation, western blot of HEK293T cells transiently transfected with ZK-circZfp609 and ZK-circZfp609-3xFlag using anti-Zfp609 antibodies revealed several proteins products (at least 4 main products) (Fig. 2.8 C). As described previously, two AUGs were normally used for translation from circZfp609, giving two protein products. Since the linear transcripts share these two AUGs with circZfp609, these two AUGs could be used for translation as well. Consequently, translation of these transcripts resulted four protein products recognized by the anti-Zfp609 antibodies.

To further examine this conclusion, western blot of HEK293T cells transfected with delSS mutants (ZK-circZfp609-delSS and ZK-circZfp609-3xFlag-delSS) was performed (Fig. 2.8 C). Notably, these delSS mutants do not express circular transcripts and therefore the anti-Zfp609 antibodies only detect protein products generated from linear transcripts. Indeed, both delSS mutants (ZK-circZfp609-delSS and ZK-circZfp609-3xFlag-delSS) gave only two proteins products. This indicated that these two protein products came from linear RNAs. The other missing products were generated from the circular RNAs (Fig. 2.8 C).

In conclusion, western blot analysis of proteins generated from overexpression vectors using the anti-Flag antibodies revealed two protein products while four protein products were recognized using the anti-Zfp609 antibodies. At the endogenous level, we only observed one protein product. Furthermore, since several transcripts were generated from the overexpression vectors, analysis of a single circle transcript using synthetic circRNA is required for a convincing conclusion about the translation of circZfp609.



**Fig. 2.12. Several proteins are produced from overexpression constructs.** A) Northern blot of inducible stable cell lines CMT93.hCG.Luc expressing circZfp609 upon doxycycline induction (CMT93-circZfp609). B) Western blot of CMT93-circZfp609 cells upon doxycycline induction detected by anti-Zfp609 antibodies. C) Scheme of ZK-circZfp609-3xFlag-First AUG construct. A stop codon in the circZKSCAN1 intron downstream of circZfp609 can be used to generate a complete ORF. D) Western blot of HEK293T cells transfected with of ZK-circZfp609-3xFlag-First AUG by anti-Flag antibodies.

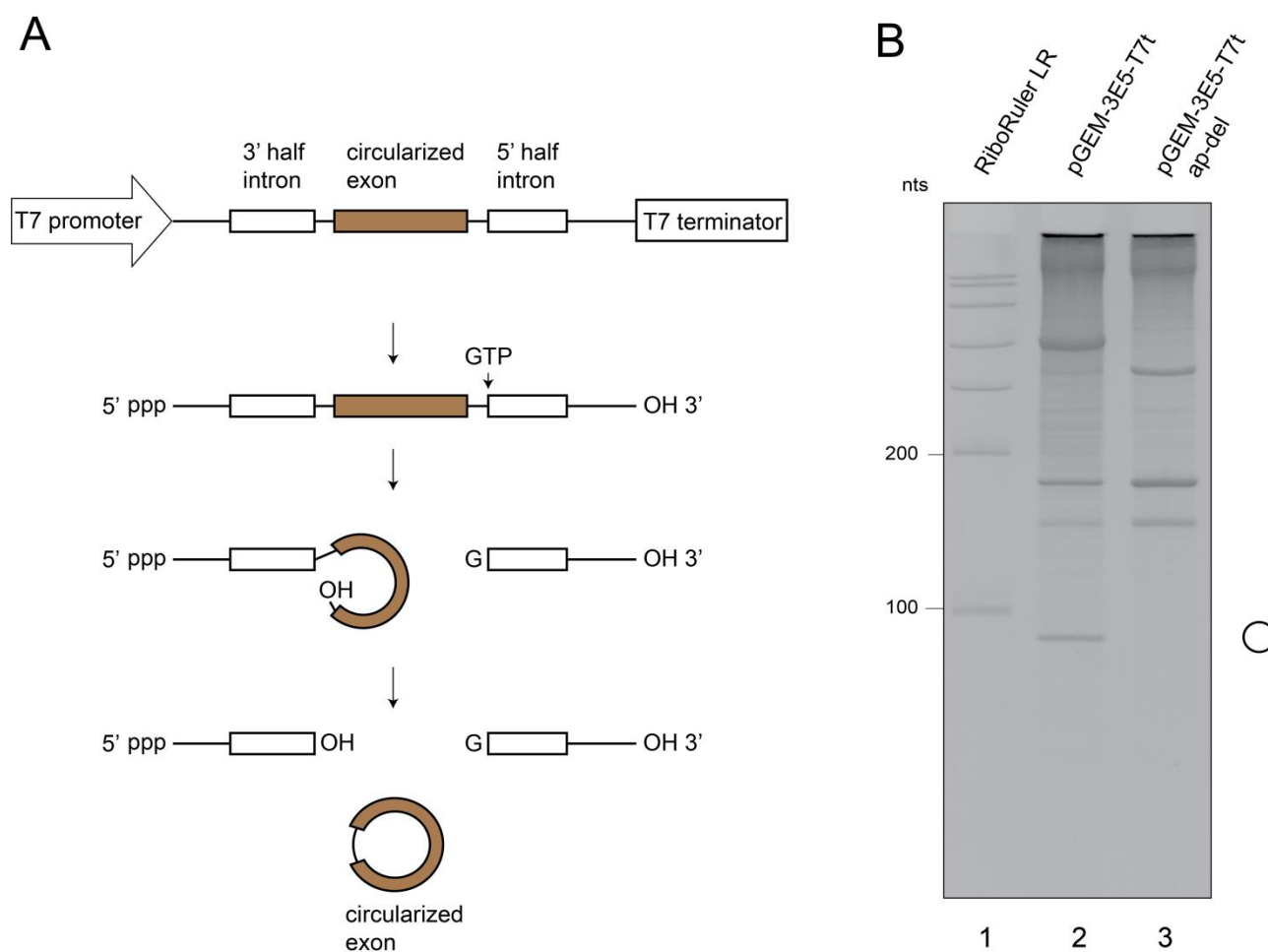
## 2.6 Synthesis of circZNF609 *in vitro*

As result from overexpression plasmids gave us several protein products recognized by the anti-Zfp609 antibodies, thus further investigation using synthetic circZNF609 are necessary to shed light on its translation mechanism. For this purpose, we performed circZNF609 synthesis *in vitro* (hereafter, we call it synthetic circZNF609) by the permuted intron-exon (PIE) method. Such studies exclude any protein bands that could be generated by the linear processing intermediates.

### 2.6.1 PIE system for *in vitro* transcription (IVT) of circRNAs

Initially, we tried to obtain synthetic circZNF609 using T7 promoter followed by RNA ligation. However, this method required several steps<sup>147</sup> and the circZNF609 can only be distinguished by running on denaturing urea polyacrylamide gel electrophoresis (Urea PAGE), which was proved to be very tricky (data not shown). To get better preparation for circZNF609, we used the PIE method. In order to test the PIE method, we followed the So Umekage's approach<sup>148</sup> (Fig. 2.13 A) using the pGEM-3E5-T7t vector. In this vector, the PIE sequence was used to produce streptavidin circular RNAs both *in vitro* and *in vivo* in *E. coli*<sup>148</sup>. The PIE sequence of the pGEM-3E5-T7t vector is formed by a 3' half intron, the streptavidin aptamer (as circularized test gene) and a 5' half intron (Fig. 2.13 A). This is the configuration necessary to achieve self-splicing and circularization<sup>148</sup>. Moreover, the PIE sequence is flanked by a T7 promoter and a T7 terminator, which allow for performing circularization *in vivo*<sup>148</sup>. The IVT of PIE method requires  $Mg^{2+}$  and guanosine nucleotides as indispensable components for splicing<sup>149,148</sup>.

To set up *in vitro* circRNA generation as described, we deleted the aptamer sequence from the vector to obtain a new plasmid, pGEM-3E5-T7t-apdel, which served as a negative control for the self-splicing reaction. We performed the IVT with an additional incubation at 55°C for 30 minutes to enhance the RNA circularization. Afterwards, the synthesized RNAs were separated onto a 6% urea PAGE (Fig. 2.13 B). EtBr staining revealed the expected bands corresponding to the aptamer circRNAs from pGEM-3E5-T7t and this band was abolished in the negative control of IVT transcribed from plasmids pGEM-3E5T7t-apdel (Fig. 2.13 B).



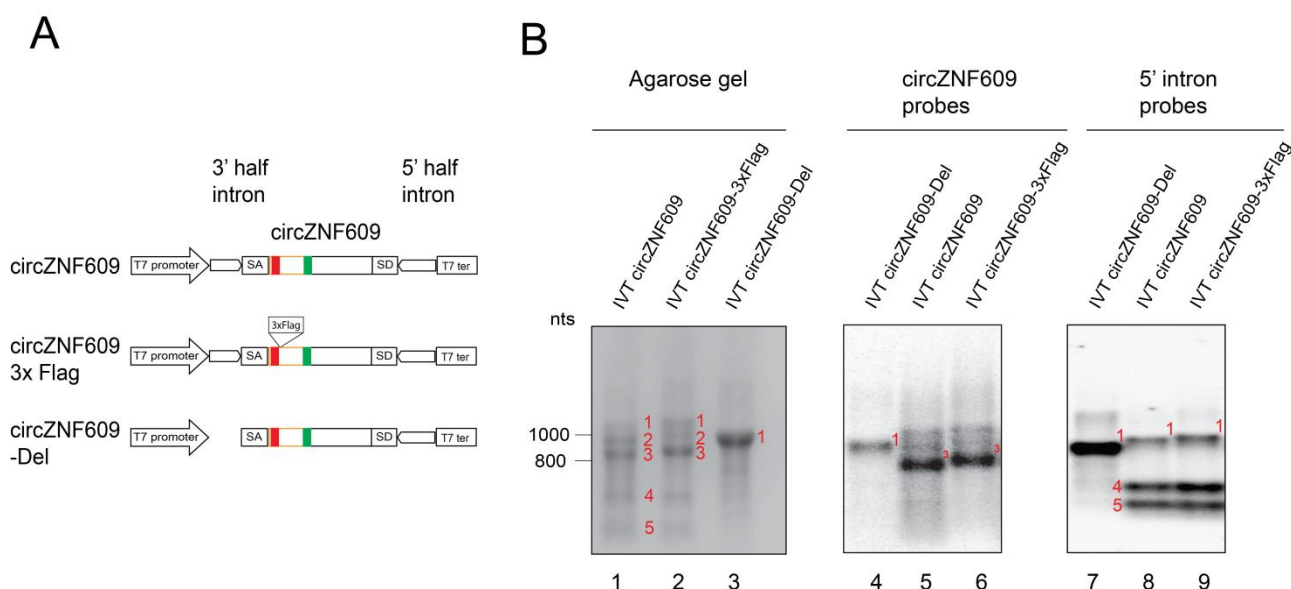
**Fig. 2.13. The PIE method for production of streptavidin circular aptamer.** A) Scheme of the pGEM-3E5-T7t vector containing all necessary sequence for PIE method<sup>148</sup>. The PIE is located between the T7 promoter sequence and the T7 terminator sequence for *in vivo* production of circRNAs in *E. coli*. The PIE sequence consists of the 3' half intron, streptavidin aptamer sequence and the 5' half intron sequence<sup>148</sup>. B) Urea PAGE of IVT from pGEM-3E5-T7t (Lane 2) and pGEM-3E5-T7t-apdel (Lane 3) plasmids.

## 2.6.2 IVT for circZNF609 synthesis

To perform IVT for production of synthetic circZNF609, the sequence of the aptamer from plasmids pGEM-3E5T7t was replaced by the sequence of circZNF609. For further analysis, several constructs were generated (Fig. 2.14 A)

After PCR amplification of the template with necessary sequence for transcription and circularization, IVT was performed. The RNA was isolated and run on 6% denaturing Urea PAGE for 4 hours and 16 hours (overnight). However, we were not able to distinguish the band corresponding to circZNF609 because of several faint bands from Urea PAGE

(Supplemental Fig. 5.13). Therefore, we ran the IVT RNA on denaturing MOPS agarose gel. EtBr staining showed 5 strong with predicted sizes (Fig. 2.14 B). Indeed, both constructs IVT circZNF609 and IVT circZNF609-3xFlag showed five bands corresponding to the full-length transcript (Fig. 2.14 B, lane 1, band 1), the linear intermediate (band 2), the presumed circular product (band 3) and the 5'-3' and the 5' half introns in the bottom (band 4 and 5) (Fig. 2.14 B, lane 1 and 2). As expected, the band corresponding to the construct circZNF609-3xFlag (Fig. 2.14 B, lane 2, band 3) was higher than its corresponding circZNF609 (Fig. 2.14 B, lane 1, band 3). Moreover, the negative control (IVT circZNF609-Del) showed one strong band representing the linear full-length transcript (band 1). These results encouraged us to proceed with northern blot validation.



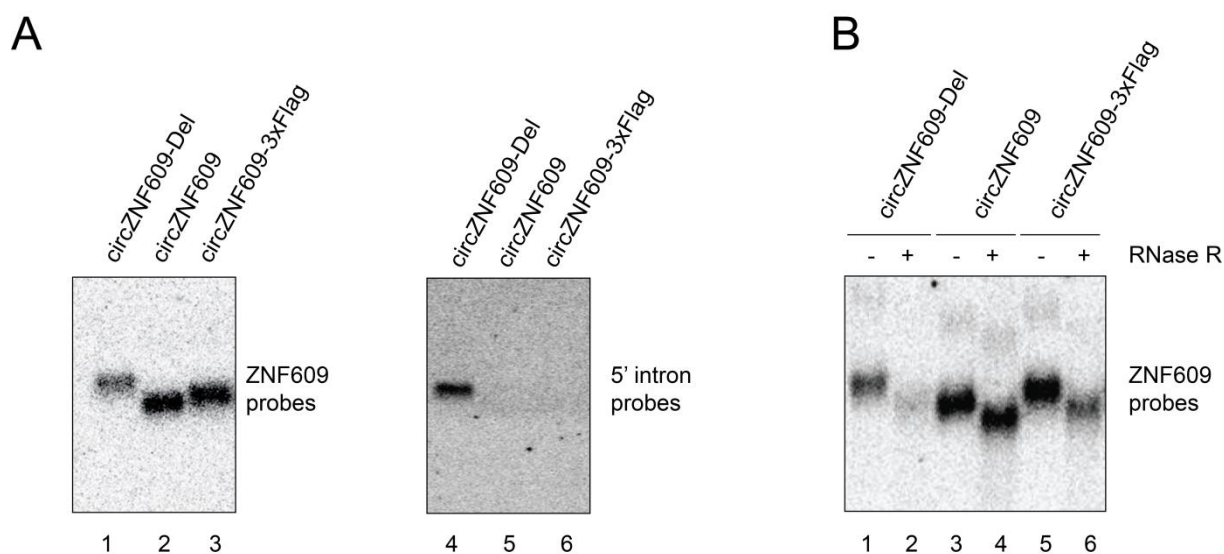
**Fig. 2.14. IVT for synthetic circZNF609 production.** A) Scheme of different constructs using the PIE methods for production of circZNF609. All constructs include SA (CUACC) and SD (GGGU) at the 3' half intron and the 5' half intron, respectively. These 9 nts are necessary for the self-splicing reaction. The clone circZNF609 del lacks the 5' half-intron and serve as a negative control. B) 1% MOPS Agarose of 10 ng of IVT RNA (lane 1, 2, 3) was performed. Northern blot analysis was detected by circZNF609 probes (lane 4, 5, 6) or 5' intron probes (lane 7, 8, 9). SA: splice acceptor, SD: splice donor.

### 2.6.3 Northern blot validation of synthetic circZNF609

To detect the circular products, a serial dilution of IVT RNA from 10 pg to 10 ng (10 pg, 100 pg, 1 ng and 10 ng) was loaded on 1% MOPS Agarose gel. At a concentration of 10 ng RNA, we obtained the best result (Fig. 2.14 B, lane 4-6). Specifically, we observed a

strong signal corresponding to the circZNF609 (Fig. 2.14 B, lane 5, band 3) and circZNF609-3xFlag (Fig. 2.14 B, lane 6, band 3) produced from IVT circZNF609 and IVT circZNF609-3xFlag, respectively. Moreover, the data showed a weak signal belonging to circZNF609-Del (Fig. 2.14 B, lane 4, band 1).

To confirm that we obtained the desired circRNAs, a second probe was designed to recognize only the 5' intron, but does not detect circZNF609. As expected, we could only observe the signal of the linear full length construct and very strong signal of circZNF609-Del (Fig. 2.14 B, lane 7-9).



**Fig 2.15. Northern blot for confirmation of isolated synthetic circZNF609.** A) 10 ng of isolated circZNF609 was loaded on MOPS Agarose gel. Northern blot was performed with ZNF609 probes (lane 1, 2, 3) or 5' intron probes (lane 4, 5, 6). B) RNase R digestion of isolated circZNF609 for confirmation the circularity of synthetic circZNF609 by northern blot.

Since after IVT, we observed five RNA products on MOPS Agarose gel, purifying the desired circZNF609 from MOPS Agarose gel is necessary for further experiments. Therefore, the bands corresponding to the circRNAs (Fig. 2.14 B, lane 1 and 2, band 3) were excised from MOPS Agarose gel and the RNA was isolated. To confirm that we obtained the desired circZNF609, 10 ng of isolated RNAs were run on MOPS Agarose for further validation by northern blot. Again, both circZNF609 and 5' intron probes were used to confirm the purified circRNAs (Fig. 2.15 A, lane 1-6). Furthermore, we used RNase R to examine the resistance of synthetic circZNF609 to exonuclease treatment. The data showed that synthetic circZNF609 and to some extent circZNF609-3xFlag is resistant to RNase R digestion (Fig. 2.15 B, lane 4 and lane 6). Finally, cDNA synthesis from isolated

RNAs and Sanger sequencing confirmed the circularity of our synthetic RNAs (data not shown).

Taken together, these data confirmed that we could develop a system for *in vitro* generation of circZNF609. Furthermore, we also established a method to purify the synthetic circZNF609 with high purity and quantity. The synthetic circZNF609 could readily be used for further analysis.

#### 2.6.4 *In vitro* translation of synthetic circZNF609

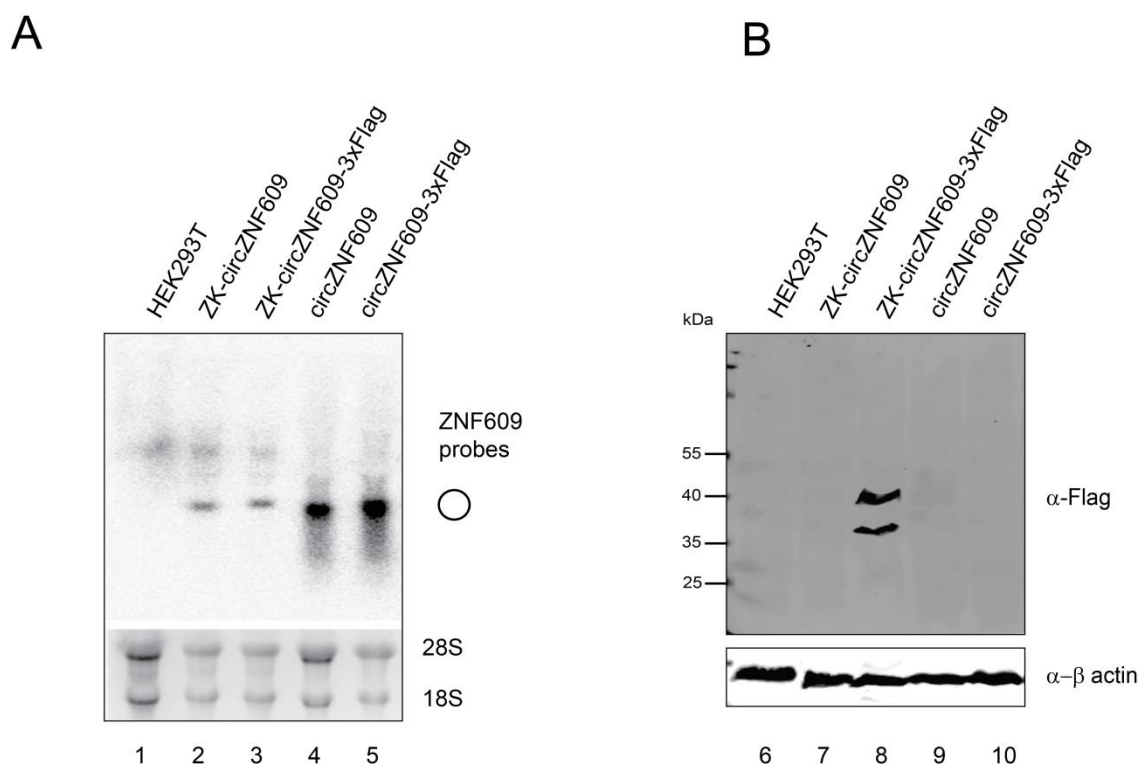
Having synthetic circZNF609 in hand, we next performed *in vitro* translation experiments using rabbit reticulocyte lysate. We could obtain the expression of the positive control from IVT GFP linear RNA (synthetic GFP) (data not shown). Of note, the synthetic GFP was designed to contain a Kozak sequence for enhancing the translation of mRNA and a 3' poly (A) tail. A 5' m<sup>7</sup>G cap was not included in the synthetic GFP.

However, we did not observe the protein expression from neither total IVT RNA nor isolated synthetic circZNF609 (data not shown). As described earlier, total IVT RNA contains the synthetic circZNF609 and four linear processing intermediates, all of which do not contain neither a 5' m<sup>7</sup>G cap nor a 3' poly (A) tail. This is because there is no 5' - m<sup>7</sup>G cap nor a 3' poly (A) tail on a circRNA. Therefore, we propose that 5' m<sup>7</sup>G cap is not required for circRNA translation but maybe other modifications are needed for efficient translation *in vitro*. Further experiments should be performed to get better understanding about translation of synthetic circZNF609 *in vitro*.

#### 2.6.5 Transfection of synthetic circZNF609 to mammalian cells

Since we did not obtain expression of synthetic circZNF609 in a cell-free translation system, analysis of synthetic circZNF609 in cells would give us some hints about translation of circRNA. Therefore, we transfected the synthetic circZNF609 into HEK293T cells. Transfection of only synthetic circZNF609 into the cells will eliminate all possible protein products generated from all linear processing intermediates. The cells were harvested 4 hours and 20 hours post transfection. Northern blot analysis showed that the synthetic RNAs were taken up by HEK293T cells 4 hours post transfection (Fig. 2.16 A). Especially, northern blot showed strong bands of synthetic circZNF609 (Fig. 2.16 B lane 4, 5) with the same size of the positive control, which is circZNF609 expressed from overexpression vector (Fig. 2.16 B lane 2, 3). However, after 20 hours, we did not observe

the expression of circRNAs by northern blot implying that the synthetic RNAs were degraded by the cells (data not shown). Western blot analysis of proteins 20 hours after transfection did not show any signal (Fig. 2.16 B) indicating that the transfected synthetic circZNF609 are not translatable.



**Fig. 2.16. Transfection of synthetic circZNF609 into HEK293T cells.** A) HEK293T was transfected with plasmids for overexpression of circZNF609 (lane 2, 3) and synthetic circZNF609 (lane 4, 5). 4 hours post transfection, RNA was harvested for northern blot analysis B) 20 hours post transfection, HEK293T cells were harvested for western blot analysis.

## 2.7 Toward understanding of circZNF609 protein function

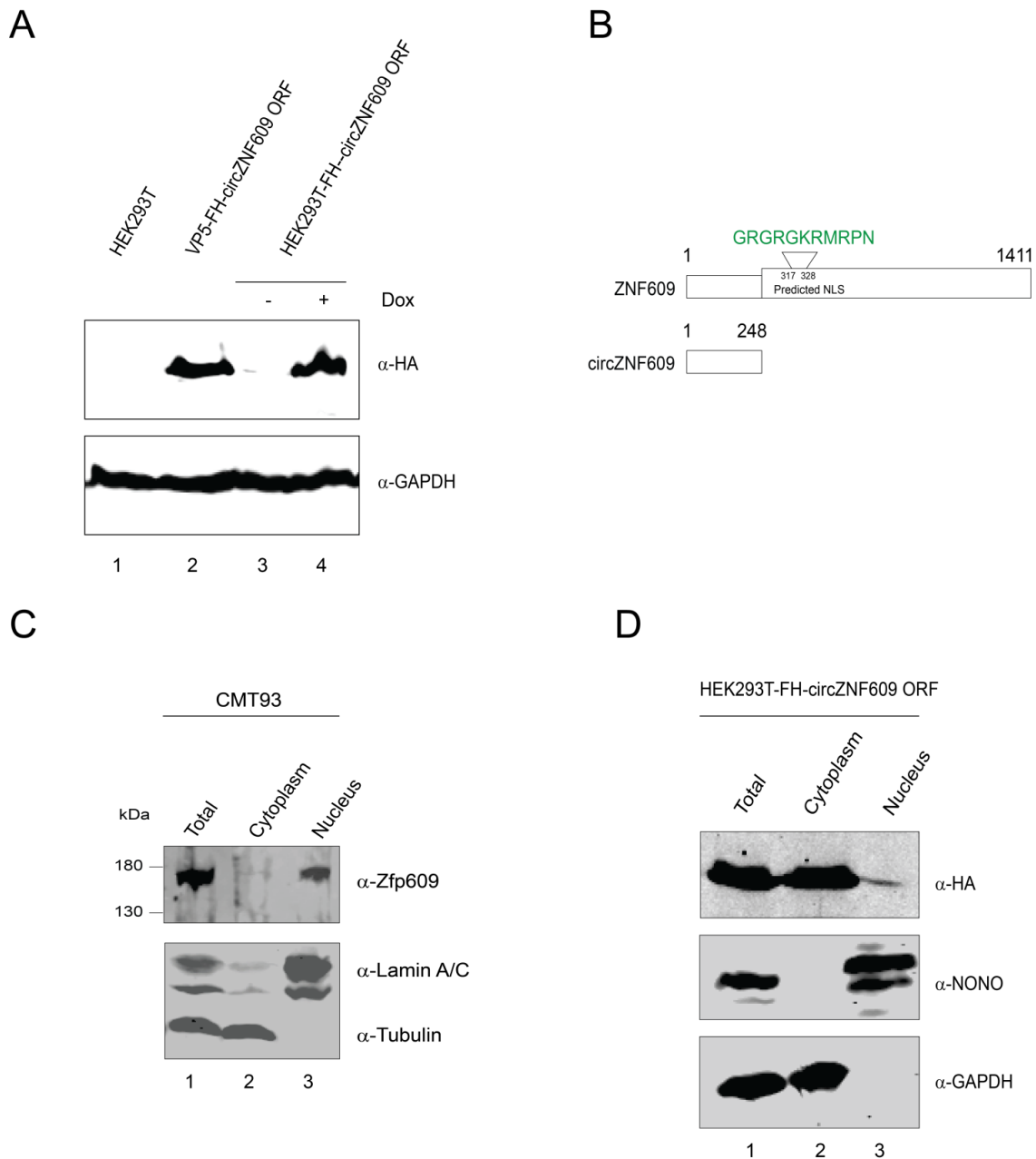
Even though *in vitro* translation and transfection of synthetic circZNF609 did not give us the expected result, further experiments to understand the function of circZNF609 protein may give us further insight into whether or not circZNF609 indeed encodes for proteins.

### 2.7.1 Generation of stable cell lines

To understand the function of proteins generated from circZNF609, the linearized form of the longest ORF of circZNF609 was N-terminally fused to Flag-HA. 3 stable cell lines were generated, including HEK293T (Fig. 2.17 A), CMT93.hCG.Luc and HCT116.hCG.Luc



(data not shown). These stable cell lines expressed the Flag-HA-circZfp609 ORF in an inducible manner upon doxycycline induction (Fig. 2. 17 A).



**Fig. 2.17. Analysis of the linearized circZNF609 ORF in HEK293T cells.** A) Western blot of HEK293T cells transiently transfected with Flag-HA-circZNF609 ORF (lane 2) and the stable cell line HEK293T expressing Flag-HA-circZNF609 ORF (HEK293T-FH- circZNF609 ORF) upon doxycycline induction (lane 3, 4). B) Prediction of the NLS in the ZNF609 full-length protein. C) Western blot of subcellular fractionation in CMT93 cells by anti-Zfp609 antibodies. Lamin A/C and Tubulin were used as nuclear and cytoplasmic markers, respectively. D) Western blot of subcellular fraction of HEK293T-FH-circZNF609 ORF after doxycycline induction using anti-HA antibodies. NONO and GAPDH served as nuclear and cytoplasmic markers, respectively.

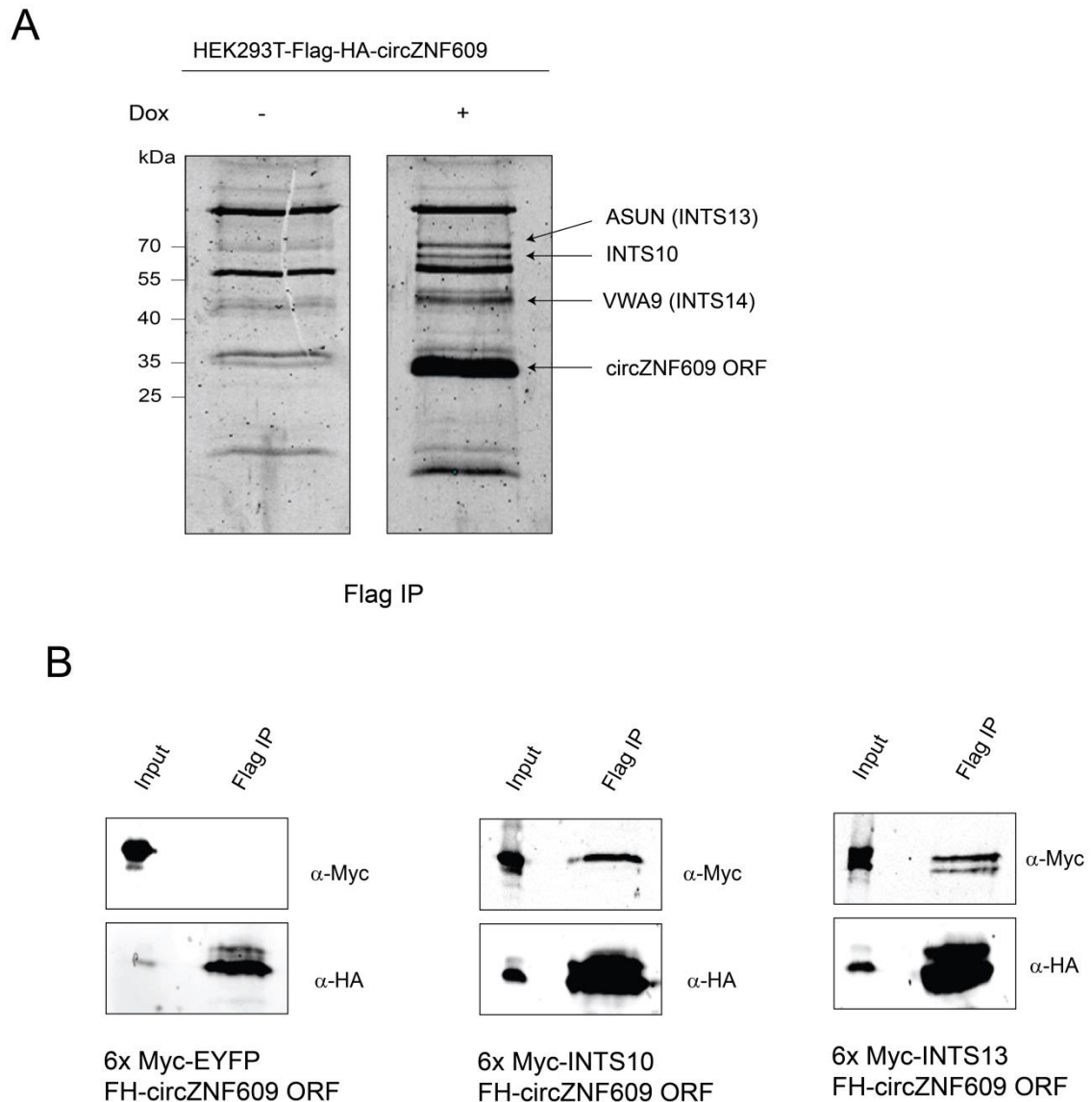
### 2.7.2 Subcellular localization of circZNF609 proteins

Analysis of subcellular localization showed that the full-length Zfp609 (and human ZNF609) protein localizes exclusively to the nucleus (Fig. 2.17 C). This is expected because Zfp609 (and human ZNF609) contain a predicted NLS (Fig. 2.17 B) and was previously reported to localize to the nucleus<sup>115</sup>. On the other hand, we observed that linearized Flag-HA-circZNF609 ORF localizes mostly to the cytoplasm (Fig. 2.17 D). Indeed, circZNF609 (and circZfp609) shares 248 aa with N-terminal of its host protein ZNF609 (and Zfp609) and circZNF609 does not contain the predicted NLS as shown for ZNF609 full-length protein (Fig. 2.17 B).

### 2.7.3 Identification of circZNF609 interaction partners

To understand further the function of circZNF609 (and circZfp609), we wanted to identify interaction partners of circZNF609 proteins. For this purpose, an anti-Flag IP from HEK293T-FH-circZNF609 ORF was performed. Mass spectrometry analysis identified several proteins of the integrator complex as interaction partners of circZNF609 proteins (Fig. 2.18 A). We could further validate these interactions by Co-IP. The data showed that INST10 and INTS13, but not EYFP binds to Flag-HA-circZNF609 ORF (Fig. 2.18 B). This is interesting because Zfp609 full-length protein was previously shown to bind to the proteins of the integrator complex<sup>115</sup> and therefore it is conceivable that the putative proteins encoded by the circZNF609 could bind some of these interactors as well. As circZNF609 protein localizes to the cytoplasm, binding of circZNF609 proteins to the integrator complex may interfere with the functions either of ZNF609 or of integrator complex in the nucleus. Since integrator complex plays important roles in processing of small nuclear RNAs (snRNAs)<sup>150</sup>, we analysed the expression of immature U1 and U2 snRNAs in HT29-sh-circZNF609 upon doxycycline induction (Supplemental Fig. 5.16). However, knockdown of circZNF609 did not have any effect in the processing of snRNAs, at least U1 and U2 snRNAs. Further analysis is required for the understanding the mechanism of circZNF609 proteins' functions.

Taken together, we found that circZNF609 proteins localize mainly to the cytoplasm while full-length ZNF609 proteins localizes exclusively to the nucleus. Furthermore, mass spectrometry analysis identified several proteins of the integrator complex as interaction partners of both circZNF609 and ZNF609 proteins suggesting an interesting mode of circZNF609's function for future investigation.



**Fig. 2.18. Identification the interaction partners of circZNF609 ORF.** A) Coomassie staining of Flag IP of the HEK293T-FH-circZNF609 ORF with and without doxycycline induction. Mass spectrometry analysis identified several proteins of the integrator complex may potentially interact with circZNF609 ORF. B) Either EYFP or INTS10 or INTS13 (all Myc-Tagged) and Flag-HA-circZNF609 ORF were co-transfected into HEK 293T cells. FLAG-IPs were performed to validate the potential binding of circZNF609 ORF with INTS10 and INTS13.

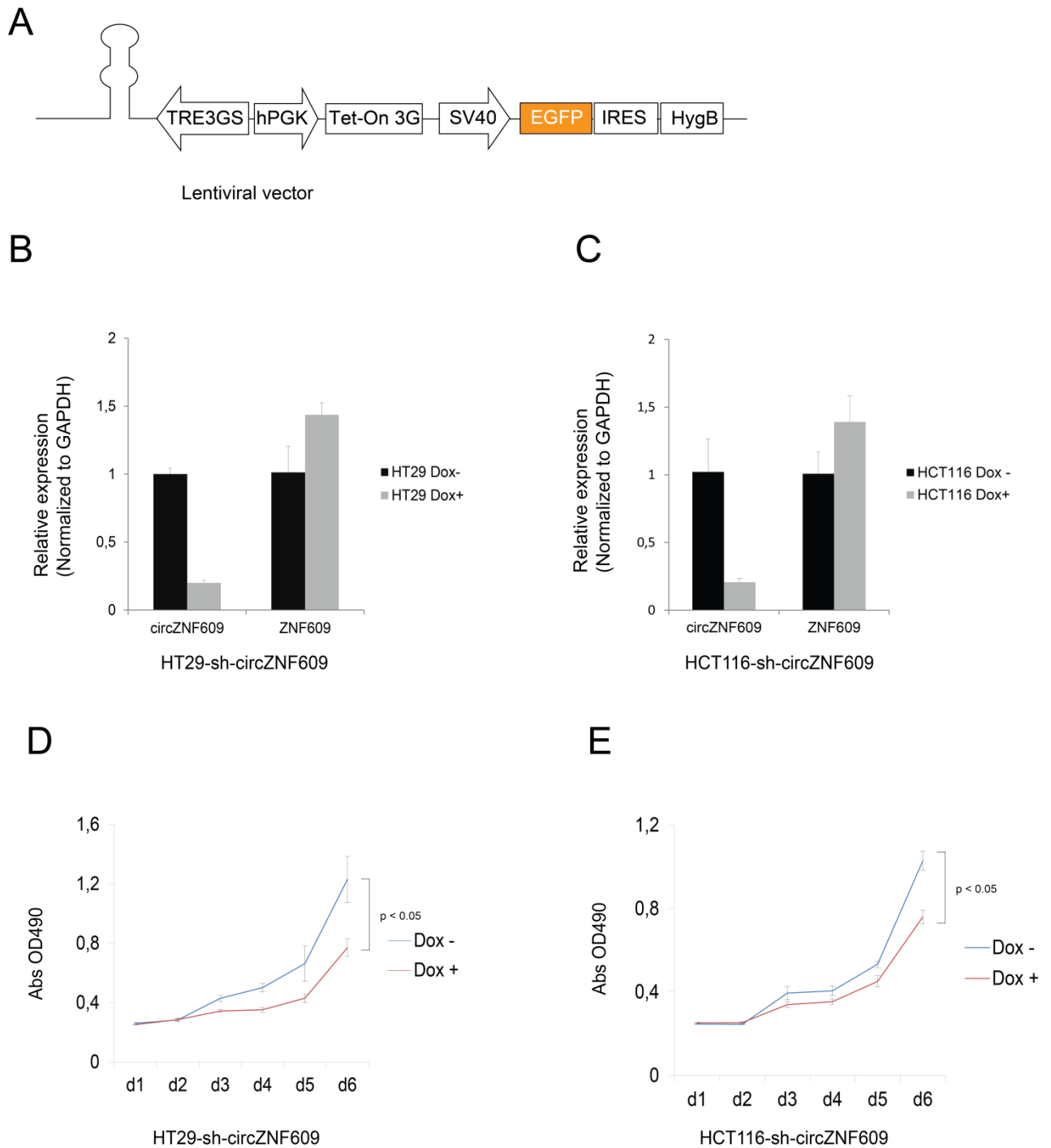
## 2.8 *In vitro* and *in vivo* validation the function of circZNF609 during colorectal cancer progression

### 2.8.1 *In vitro* proliferation assay

From RNAseq analysis of our mouse model, circZNF609 is upregulated in primary tumor as well as liver metastasis suggesting that circZNF609 may play a role in development of colorectal cancer. To investigate the impact of circZNF609 on colon cancer development *in vivo*, we generated different stable cell lines, including HT29-sh-circZNF609, HCT116-sh-circZNF609 and CMT93-sh-circZfp609, which itself can produce siRNAs to knock down ciZNF609 upon doxycycline induction. HT29-sh-circZNF609 and HCT116-sh-circZNF609 cell lines stably expressing shRNA targeting circZNF609 were generated using lentivirus (Fig. 2.19 A) while CMT93-sh-circZfp609 stable cell lines were obtained using transient transfection of construct described in Fig. 2.4 C.

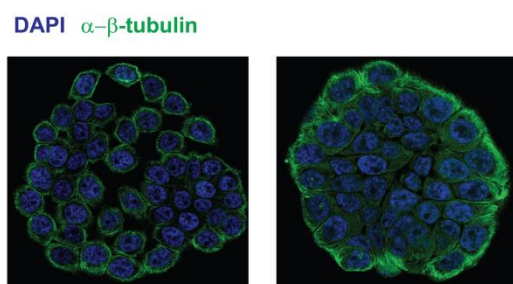
To confirm the expression of siRNA in an inducible manner, northern blot was performed three days after doxycycline induction. Northern blot revealed that the desired siRNAs were induced significantly upon induction with doxycycline in all cell lines (data not shown). Subsequently, we observed a significant knockdown of circZfp609 without affecting the expression of Zfp609 mRNA in CMT93-sh-circZfp609 cell line (Fig. 2.21 A). Moreover, we could also specifically target circZNF609 in both HT29-sh-circZNF609 and HCT116-sh-circZNF609 polyclonal cell lines (Fig. 2.19 B, C).

To address the effect of circZNF609 knockdown on the proliferation, XTT assays were performed. The data showed that knockdown of circZNF609 upon doxycycline induction decreases cell proliferation of both HT29-sh-circZNF609 and HCT116-sh-circZNF609 cells (Fig. 2.19 D, E). Further investigation by immunofluorescent (IF) analysis of HT29-sh-circZNF609 cells four days treated with doxycycline revealed different cell behavior between doxycycline-treated and nontreated ones (Fig. 2.20 A). In this experiment, we seeded a single cell on coverslip and four days after induction, the cells were fixed and stained with anti- $\beta$ -tubulin. Especially, the data showed that upon induction of doxycycline, the cells tend to stick together compared to the non-treated ones. This implied that the decrease in proliferation could be result of a defect in cell division (Fig. 2.20 A). In the future, it would be interesting to study the effect of circZNF609 knockdown on cell-cell interaction and migration *in vitro*.



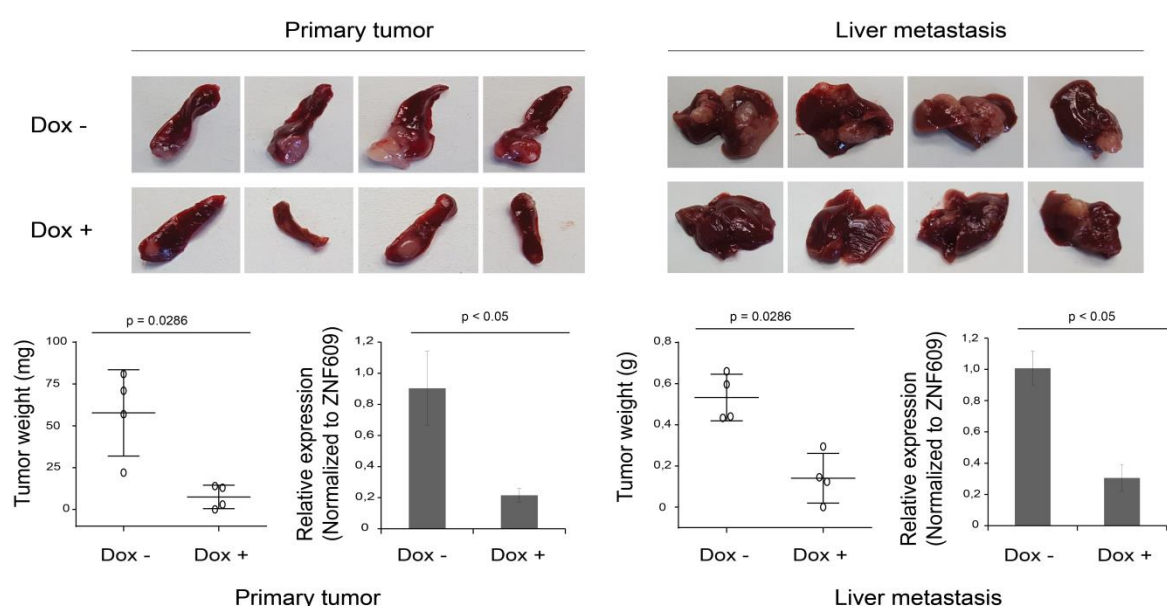
**Fig. 2.19. Knockdown of circZNF609 decreases cell proliferation.** A) Part of lentiviral vector for knockdown of circZNF609 via viral transduction. B) C) Relative expression of circZNF609 and ZNF609 in the stable cell line HT29-sh-circZNF609 (B) and HCT116-sh-circZNF609 (C) upon doxycycline induction validated by qPCR. The relative expression of circZNF609 was normalized to the expression of GAPDH. D) E) Knockdown of circZNF609 decrease cell proliferation in HT29-sh-circZNF609 (D) and in HCT116-sh-circZNF609 (E). Dox: Doxycycline

A



HT29-sh-circZNF609

B



**Fig. 2.20. Knockdown of circZNF609 decreases colon cancer progression and metastasis *in vivo*.** A) IF of the HT29-sh-circZNF609 colony four days after induction with doxycycline. The cells were stained with DAPI and anti- $\beta$ -tubulin antibodies. B) HT29-sh-circZNF609 cell lines were injected into spleen of the SCID mice. Spleens contain primary tumors and livers with metastases were imaged for macro analysis. The according tumors were resected and measured by tumor weight. Relative expression of circZNF609 of RNA isolated from primary tumors and liver metastasis by qPCR. The relative expression was normalized to ZNF609. Dox: Doxycycline.

## 2.8.2 *In vivo* xenograft mouse model

### 2.8.2.1 HT29/SCID

Since depletion of circZNF609 in HT29-sh-circZNF609 cells upon doxycycline induction led to decrease in cell proliferation *in vitro*, we wanted to test the effect of circZNF609 knockdown *in vivo*. Therefore, 500.000 HT29-sh-circZNF609 cells were injected into

spleen of SCID mice. Depending on the signal of luciferase, mice were sacrificed accordingly. The data showed that upon induction with doxycycline, less primary tumor development as well as less liver metastases compared to the control group was observed (Fig. 2.20 B). The RNA from those tumors was collected for qPCR analysis. qPCR analysis showed a clear decrease in the expression of circZNF609 while the expression of ZNF609 was not changed (Fig. 2.20 B).

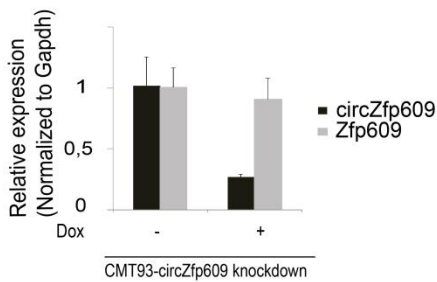
#### 2.8.2.2 CMT93/C57BL/6 for knockdown of circZfp609

As the phenotype obtained from our knockdown experiments could also be due to off-target effects, we attempted to design a second siRNA to confirm our data. However, the second siRNA should also target the junction of the circZNF609 since this feature is unique to circRNA. However, designing such an siRNA which does not target its cognate host gene is difficult. To overcome this problem, we searched for potential off-targets of our siRNA within the human and mouse transcriptome. We found that the top potential targets of this siRNA in human and mouse are different (Table 3). Therefore, if we get the same phenotype of circZNF609 knockdown using human and murine cell lines, it is unlikely because of the off-target effects.

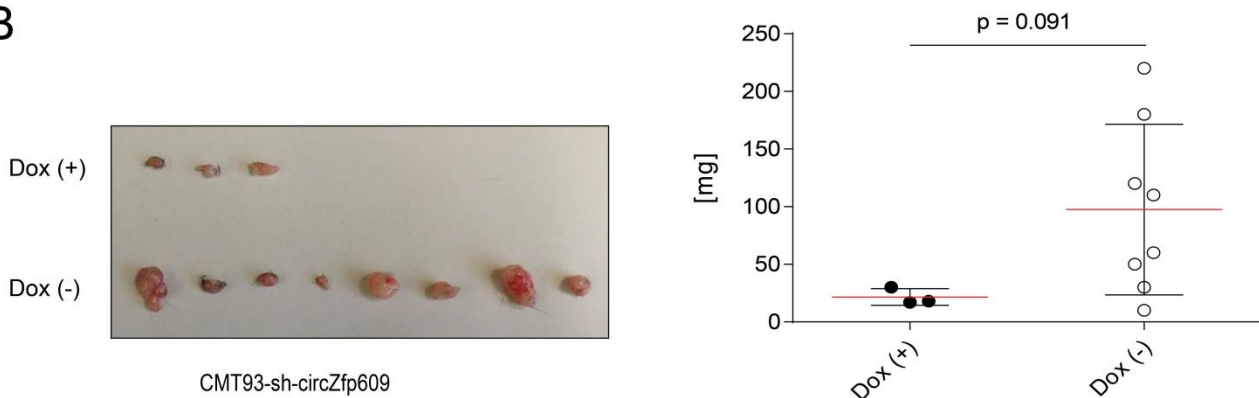
Analysis of human circZNF609 and murine circZfp609, we found that the designed siRNA can target the murine circZfp609 as well. Therefore, we decided to validate the effect of circZfp609 knockdown in the murine CMT93.hCG.Luc cell lines (Fig. 2.21 A). To study the function of circZfp609 *in vivo*, 500.000 CMT93-sh-circZfp609 cells were subcutaneously injected into C57BL/6 mice. The data showed that upon induction with doxycycline, the tumors are smaller (Fig. 2.21 B, C). However, more mice are needed for statistically significant data. From what we observed, this *in vivo* data is consistent with the data obtained from the HT29-sh-circZNF609/SCID mouse model.

Taken together, the result from circZNF609 knockdown using human colorectal cancer HT29-sh-circZNF609 cell line transplanted into SCID mice supports our cell proliferation experiments *in vitro* suggesting that circZNF609 indeed plays a role during colon cancer progression. Furthermore, we obtained the same data when we knocked down murine circZfp609 in CMT93-sh-circZfp609 cell line transplanted into C57BL/6 mice. Therefore, we conclude that the phenotype obtained from knockdown experiments is not because of off-target effects.

A



B



**Fig. 2.21. Knockdown of circZfp609 decreases colon cancer progression.** A) Relative expression of circZfp609 and Zfp609 in the stable cell line CMT93-sh-circZfp609 upon doxycycline induction validated by qPCR. The relative expression of circZfp609 was normalized to the expression of GAPDH. B) *In vivo* data of tumor resected from the mice subcutaneously injected with CMT93-sh-circZfp609 upon doxycycline induction. The tumor development was calculated by measuring the tumor weight. Dox: Doxycycline.

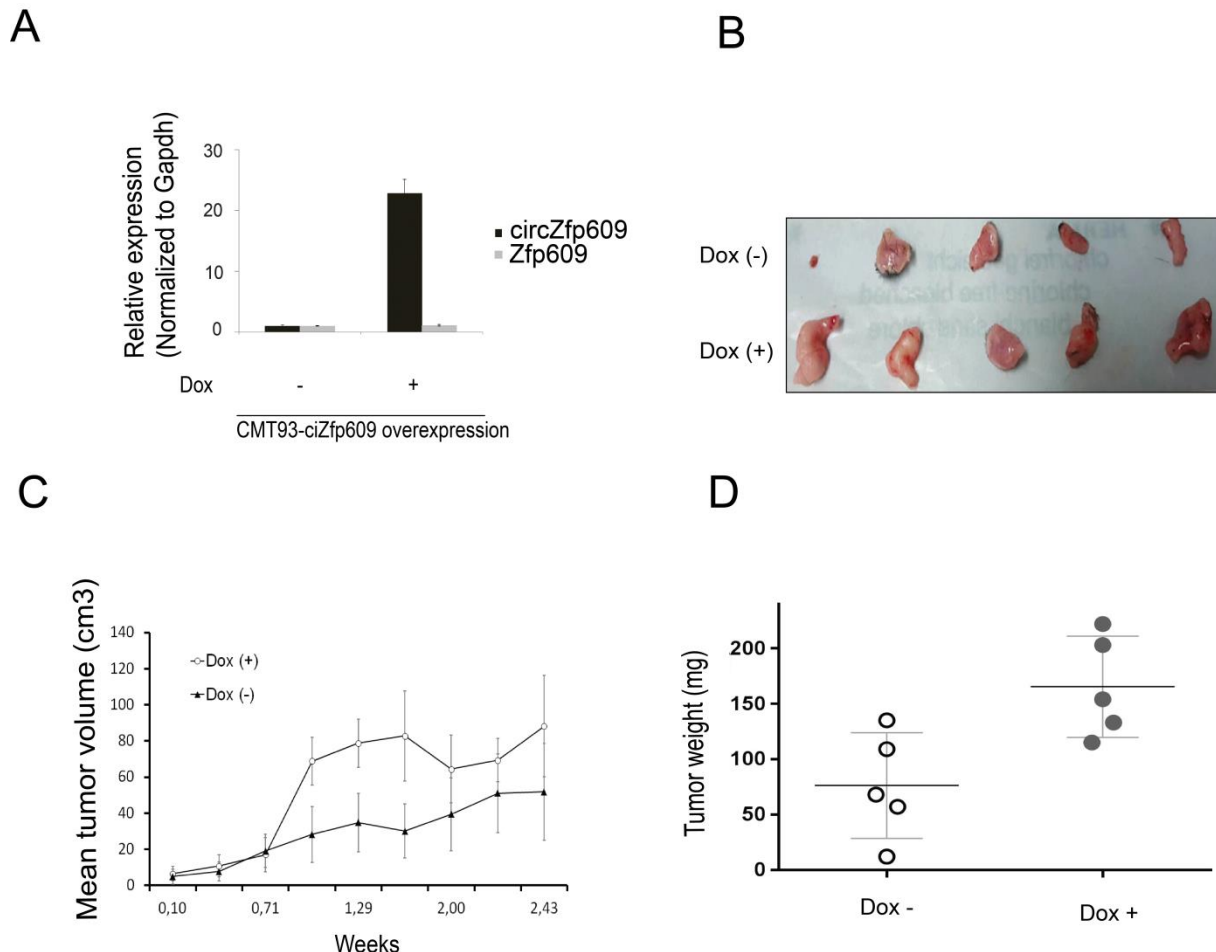
### 2.8.2.3 CMT93/C57BL/6 for overexpression of circZfp609

To further confirm the phenotype obtained from circZNF609 knockdown, we wanted to test the effect of circZNF609 overexpression *in vivo*. Therefore, we generated an inducible stable cell line CMT93.hCG.Luc expressing circZfp609 (CMT93-circZfp609) upon doxycycline induction (Fig. 2.12 B). As expected, the qPCR data show upregulation of circZfp609 upon doxycycline induction (Fig. 2.22 A).

Next, we performed *in vivo* experiments by injecting 500.000 cells of the CMT93-circZfp609 stable cell line subcutaneously into C57BL/6 mice. The data showed that upon induction with doxycycline, the mice developed a bigger tumor size compared to the non-induction control (Fig. 2.22 B). Consequently, this resulted in bigger tumor volume (Fig. 2.22 C) as well as tumor weight (Fig. 2.22 D) upon doxycycline induction. Since the data



obtained from overexpression experiments is opposite to the data obtained from knockdown experiment, this further supports our phenotype we observed from previous experiments.



**Fig. 2.22. Overexpression of circZfp609 promotes colon cancer development.** A) CMT93-circZfp609 stable cell line was induced by doxycycline for 3 days. Afterward, the cells were harvested for qPCR analysis. B) C) D) *In vivo* data from the mice overexpressing circZfp609 upon doxycycline induction (n=10). Induction of circZfp609 by doxycycline increases tumor volume (C) and tumor weight (D). Dox: Doxycycline.

In conclusion, *in vitro* and *in vivo* experiments confirmed that circZNF609 functions as oncogene to promote the progression of colorectal cancer. Further experiments are needed to study if circZNF609 proteins have any function in colon cancer development. Moreover, it is not clear whether circZNF609 functions as non-coding or coding RNA. Therefore, it would be very interesting to investigate this aspect in the near future.

### 3 Discussion

CircRNAs are becoming the new regulatory class of non-coding RNAs. Several studies have revealed that circRNAs are differentially expressed in several tissues and diseases. Consequently, some of circRNAs were shown to play important roles in several diseases such as cancer, cardiovascular disease and neurological disease. So far, the molecular mechanisms for most of circRNAs are related to miRNA sponges.

In this study, we found that circZNF609 is upregulated during colorectal cancer progression and metastasis. *In vitro* and *in vivo* experiment implied that circZNF609 can function as oncogene. Interestingly, we found that circZNF609 could potentially encode for proteins and these proteins share the AUGs with its host mRNA, ZNF609. Using Flag reporter constructs and mutagenesis, we could show that the translation of circZNF609 is possible. Furthermore, mass spectrometry analysis identified several proteins of the integrator complex as interaction partners of circZNF609 proteins. We found that circZNF609 proteins localize to the cytoplasm whereas ZNF609 proteins localize exclusively to the nucleus. Since both circZNF609 and ZNF609 bind to the integrator complex, further analysis of these interactions would shed lights on how circZNF609 functions.

#### 3.1 Overexpression of circZfp609

Several labs attempted to overexpress circRNAs in mammalian cells. The first overexpression construct was produced from the Kjems lab where they used base pairing between the long inverted sequences of flanking intron for producing circRNA, as shown for ciRS-7<sup>70</sup>. Furthermore, as intron pairing is a cis element for biogenesis of circRNAs, the Willusz lab showed that circRNAs can successfully be generated using several introns, including intron flanking circZKSCAN1 or circLaccase 2<sup>52</sup>. Importantly, these plasmids were shown to be more efficient in producing circRNAs compared to the method developed by the Kjems lab<sup>52</sup>. Another study used the overexpression construct for circRNA production with its endogenous flanking intron as shown for circMbl<sup>45</sup>. Others used the type 1 intron self-splicing of bacteriophage T4 td gene for overexpression as shown for circFoxo3<sup>105</sup>. However, compared to the above methods, this is not the ideal method because there are extra nucleotides included in the overexpressed circRNAs which could be problematic for further functional studies.

In this study, the circZfp609 is produced using the flanking intron of circZKSCAN1, which was developed by the Willusz lab. Consistent with the previous conclusion<sup>52</sup>, we also observed that using the circZKSCAN1 intron is more efficient in producing circRNAs compared to methods developed by the Kjems lab<sup>70</sup> (Fig. 2.5 B). The fidelity of overexpressed circZfp609 was confirmed by comparing to the expression of endogenous circZfp609 (Fig. 2.5 B). Further validation including Sanger sequencing and RNase R treatment proved that murine circZfp609 is expressed in human HEK293T cells. These results indicated that circZfp609 is successfully generated using chimeric circZfp609-circZKSCAN1 intron.

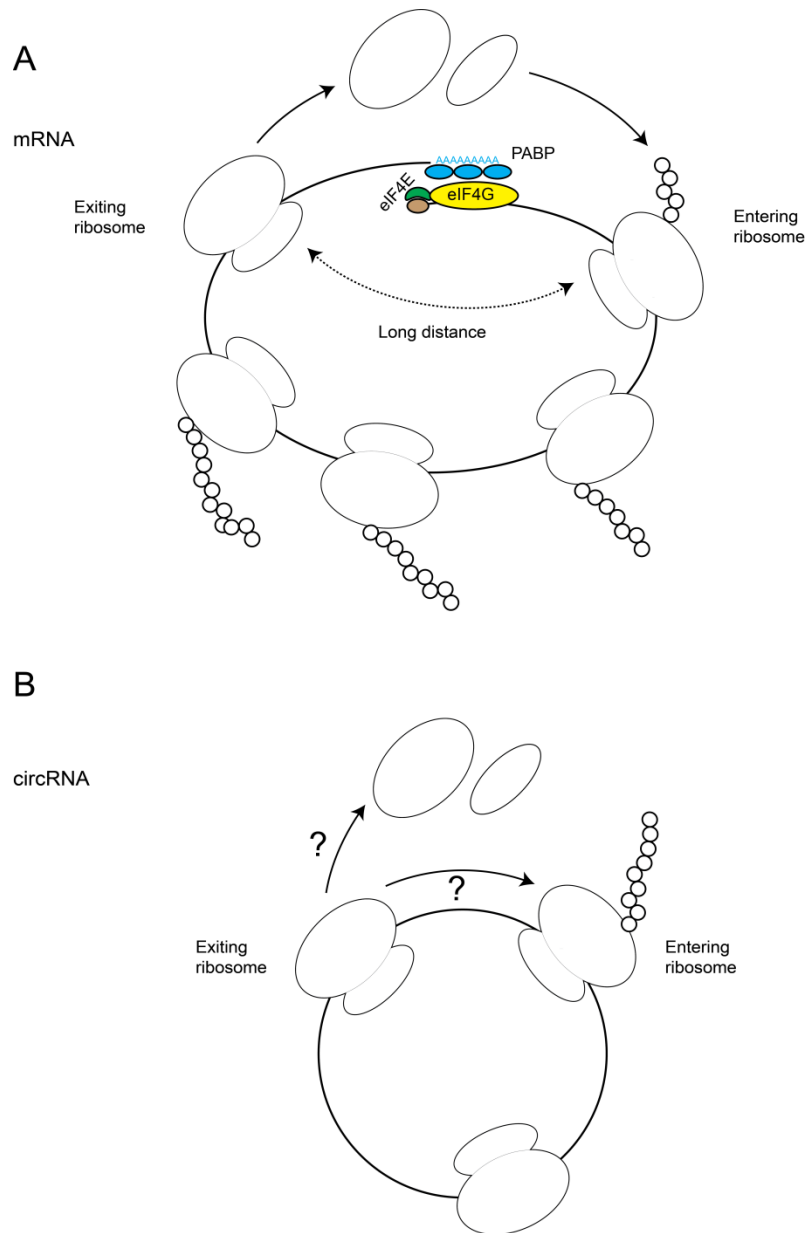
Of note, even though the overexpression of circZfp609 using the ZKSCAN1 intron worked very well, we still observed several linear RNA products that are produced from this overexpression construct (Fig. 2.5 B). This may interfere with the phenotype of circZfp609. Moreover, due to the longer half-lives of circRNAs compared to their linear forms, we observed that the stable cell lines CMT93-circZfp609 show more circRNAs than linear RNAs upon induction of doxycycline (Fig. 2.12 B). This is in contrast to the transiently expressed transcripts from the CMV promoter, where we observed more linear processing intermediates than circRNAs (Fig. 2.5 B). Specifically, the transient expression from CMV promoter was analysed one day post transfection while inducible expression by doxycycline was analysed three days after induction. Therefore, several linear processing intermediates, which were highly expressed after transcription from CMV promoter, were most digested after three days of induction by the cells.

Several studies showed that circRNAs can have non-coding functions<sup>33</sup> as scaffolds for RBPs<sup>104,105,108</sup> or miRNAs<sup>151</sup> and served as so-called “sponges”. However, one should keep in mind that there is precursor linear RNA generated together with circRNAs. Since the sequence of these precursor RNAs contain entire sequence of circRNAs, such scaffold or miRNA sponges should be investigated carefully to be convincing because the linear forms could serve as sponges as well.

## 3.2 Translation of circZfp609

It is very interesting to observe that translation of circRNA, in this case circZfp609, occurred *in vivo* in mammalian cells when fused to a 3xFlag epitope tag. We observed at least 3 AUGs from circZfp609 can be used to initiate the translation of circZfp609. In wildtype constructs, only 2 AUGs were used (Fig. 2.5 C). The third AUG was used only

when the first AUG was mutated to ACG, much better than when the second AUG was mutated (Fig. 2.6 B).



**Fig. 3.1. Topological distance between entering ribosome and exiting ribosome on mRNA and circRNA.** A) On mRNA, the binding of eIF4E, eIF4G and PABP forms a highly stable, a closed-loop structure of circular mRNA–protein complex<sup>152,153</sup>. Even though there is existence of a circle structure, ribosome enters at a 5' cap and exits near the polyA tail in a far topological distance. B) On circRNA, such initiation proteins are not yet known. However, the entering ribosome and exiting ribosome are presumably happened at very close topological distance.

As suggested in the introduction (see 1.4.2), in a circRNA, the beginning of the ORF is topologically close to the end of the ORF (Fig. 3.1). This means that initiation and

termination will presumably occur in close proximity and the two processes might positively or negatively influence each other. This could explain why we only observed the usage of the third AUG when the first AUG is mutated (Fig. 2.6 B). Moreover, we observed the usage of all 3 AUGs from circZfp609 when we lengthened the UTR of circZfp609 by adding 49 nts unrelated sequence (in this case the  $\beta$ -globin 5' UTR) just in front of the first AUG. Adding these 49 nts might increase the proximity between initiation and termination and therefore, the third AUG was used for translation.

Because cellular mRNAs have a m<sup>7</sup>G cap at the 5' end of mRNA, the translation of these RNAs are followed the cap-dependent scanning mechanism<sup>152</sup>. However, when the cap-dependent scanning initiation is inhibited, some cellular mRNAs can alternatively be translated by cap-independent mechanisms, for examples initiation via IRES<sup>154</sup>. Since circRNAs do not contain m<sup>7</sup>G cap, translation on circRNAs is supposedly initiated through cap-independent mechanism mediated by IRES sequence. However, we did not observe IRES sequence as described for other circRNAs, such as circMbl or circFBXW7, on circZfp609. As IRES determines the sites where the ribosomes enter to the circles, we speculate that on circZfp609, the ribosomes can initiate translation of the circRNA-embedded ORF at several place. This is consistent with our observation that several AUGs, at least seven AUGs, can be used for initiation of circZfp609. Indeed, we observed that four AUGs from the exogenous sequence, in this case 3xFlag tag, can be used for initiation of translation suggesting that adding some sequence can affects circRNA translation. Such observation was reported previously that the Flag sequence itself can drive translation without any specific IRES<sup>96</sup>. Taken this evidence together, we suggest that the scanning mechanism of mRNA translation is not applicable to circZfp609, and potentially even other circRNAs.

In theory, ribosomes can go around the circle several times, if it is not properly terminated. It has been observed for an infinite GFP ORF<sup>94</sup>, ribosomes can "travel" several times, more than 10 times. As mentioned in the result, we observed that in our circZfp609 construct Del121-AU, the four AUGs from 3xFlag tag can be used for translation initiation giving four ORFs with the longest ORF is 873 nts (Fig 2.11 D). However, the Del121-AU circZfp609 mutant is a 820 nts circRNA suggesting that ribosomes traveled 1,065 (873/820) times of the circle (Fig 2.11 D). Therefore, we proposed (at least for circZfp609) that the ribosomes can recognize and initiate translation from several AUGs and can travel

on a circRNA more than one circle. Such rolling circle translation was previously reported in both prokaryote and eukaryote<sup>95,96</sup>.

How ribosomes can be recycled on a circRNA? In bacteria, after completing translation, the ribosomes are disassembled for the next round of translation through the recycling process<sup>155,156</sup>. Moreover, several studies proposed that the ribosome recycling factor (RRF), elongation factor (EF-G) and initiation factor IF3 are involved recycling in prokaryotes<sup>157</sup>. Even though recycling of ribosomes has been well studied for mRNA in prokaryotes, the RRF homologue has not been found in eukaryotes suggesting that the recycling may require different factors<sup>158</sup>. Indeed, it was shown that initiation factors eIF3, eIF1, eIF1A<sup>158</sup> and ABCE1<sup>159,160,161</sup> can promote recycling of eukaryotic ribosomes. Since ribosomes can translate several times on a circle, for example, on the mentioned infinite GFP ORF<sup>94</sup>, it is not clear whether or not ribosomes are disassembled on a circRNA (Fig. 3.1). Moreover, if ribosomes are recycled, do they use the same factors as mRNA? In some cases, after termination, ribosomes do not undergo complete recycling and reinitiation was observed<sup>162,163</sup>. Therefore, in the future, it is interesting to examine if ribosomes reinitiate a new round of translation without recycling on a circRNA.

### 3.3 CircRNA synthesis *in vitro*

During the course of this study, several efforts have been made to generate synthetic circRNAs *in vitro*. One of the approaches used ligation of transcribed RNA. It works very well for several small circRNAs but it is more challenging for long circRNAs (unpublished data from the Meister lab). Moreover, circRNAs have slower mobility comparing to the linear RNAs with the same sizes on Urea PAGE. At first, we tried to run circZNF609 on Urea PAGE gel with percentage ranging from 4-6%. However, we could not obtain good resolution for circZNF609 (Supplemental Fig. 5.13). Moreover, the yield of circZfp609 we could get from UREA PAGE was very low. Another problem is that the running mobility of circRNA on MOPS Agarose gel is different from Urea PAGE. Specifically, circRNAs have the same mobility with linear RNA at the same sizes on MOPS Agarose. Therefore, such a ligation-based approach is not distinguishable on MOPS Agarose gel.

Later on, we switched to *in vitro* self-splicing group I introns from genes coding for thymidylate synthase (td) of bacteriophage T4<sup>164,165</sup> using PIE methods<sup>149,166,167,148</sup>. This is effective method for producing circRNAs and was recently reported in several studies<sup>111,148</sup>. However, the splice sites from Td gene of bacteria, which contains 4 nts at

the 3' intron and 5 nts at the 5' intron, are necessary for producing circRNAs<sup>148</sup>. Therefore, the resulting synthetic circRNAs contains 9nts from the intron. Such extra nts should be taken into account because of several reasons. First, the junction of circRNAs is the unique sequence that distinguishes circRNAs from their cognate mRNAs. We would speculate that the junction of circRNAs may contain binding sites for miRNAs or RBPs, which are different from their host mRNAs. When engineering such synthetic circRNA using *in vitro* self-splicing, adding 9 nts extra to the junction of circRNA may disrupt these sites. Secondly, adding 9nts may change secondary structure of circRNAs. Since circRNAs are considered as non-coding RNAs, such changes may affect binding to other miRNAs or RBPs and eventually their function. Thirdly, transfecting circRNAs generated by *in vitro* self-splicing could stimulate innate immune response as shown recently<sup>111</sup>.

### 3.4 Transfection of circRNAs into mammalian cells

Northern blot revealed that circZNF609 is taken up from HEK293T cells 4 hours post transfection. After 20 hours, circRNAs were mostly degraded. Such degradation pathways of circRNAs are not yet reported and it is interesting to understand which nucleases, presumably endonucleases, degrade these circRNAs.

To investigate the translation potential of transfected synthetic circZNF609, western blot was performed with HEK293T cells 20 hours post transfection. However, we did not observe any band belonging to the expected proteins. There are several reasons which can explain this result. First, synthetic circZNF609 may change its secondary structure. This may affect its potential to recruit ribosomes for translation. Second, the translation mechanism of circRNAs is not yet characterized and the synthetic circZNF609 itself may not fulfill all requirements for protein translation in mammalian cells. For examples, one study suggested that m6A is important for translation of a subset of circRNAs. Indeed, m6A promote initiation of protein translation from circRNAs in human cells by recruiting initiation factor eIF4G2 and m6A reader YTHDF3<sup>84</sup>. If it is also true for circZNF609, it should be m6A methylated *in vitro* before transfecting to the cells so that ribosomes can be recruited afterwards. Third, synthetic circZNF609 was already degraded 20 hours post transfection. This means that the cells recognize the synthetic circZNF609 as unwanted products and transfer those products immediately to the degradation pathway.

### 3.5 Effect of circZNF609 on colorectal cancer progression *in vitro* and *in vivo*

As shown by MTT, cells grew slower when we knocked down circZNF609 upon doxycycline induction. Of note, doxycycline does not give any effect to the cells because treatment of doxycycline to HCT116 cell does not lead to any changes in cell growth (Supplemental Fig. 5.15). Since only one siRNA was used in this experiment, the observed phenotype could still be an off-target effect. However, as siRNA targeting circRNA is designed only at the junction of circRNAs, there are not many choices for designing such siRNAs. One solution would be moving the siRNA a few nts way from each other. However, in any case, the siRNA should not target the cognate mRNAs.

Since this siRNA targets both human circZNF609 and murine circZfp609, we searched for potential off-targets in the human and mouse transcriptomes. We observed that this siRNA target different transcripts in human and mouse. Moreover, the *in vivo* data observed from HT29/SCID and CMT93/C57BL/6 mouse model gave us the similar phenotype. Since it is unlikely that different off-targets lead to the same phenotype, this could strengthen our data about the on-target effects of circZNF609 knockdown on tumor development.

To further support our data, we generated the stable cell line CMT93-circZfp609 for overexpression of circZfp609 upon doxycycline induction. In this case, we observed an increase in tumor development *in vivo*. This is consistent with the RNAseq data from our mouse model, which showed that circZfp609 is upregulated in primary tumor and liver metastasis. Thus, the data from circZfp609 overexpression was opposite to the data obtained from circZfp609 knockdown, confirming the potential roles of circZfp609 in tumor progression.

Even though these data sets are convincing, the rescue experiments are necessary to fully support our data. However, such rescue experiments are rather difficult and tricky. We have tried to rescue the phenotype of circZfp609 knockdown but several things should be taken into account. First, we do not know if the function of circZfp609 is from its non-coding or coding potential. Evidence from our data suggests that circZfp609 may encode for proteins using overexpression constructs. However, we do not know if the proteins from circZfp609 are produced or if produced, are really abundant in the cells. If circZfp609 encodes for proteins, the designing of resistant construct for rescue experiment is straightforward by following the silent mutation rules. However, if circZfp609 does not



encode for protein and function as non-coding RNA, the design of the resistant construct should follow other rules to maintain the structure, and therefore the function, of circZfp609.

To overcome this, we tried to identify the cis sequence in the endogenous intron of circZNF609 that is important for biogenesis of circZNF609. If we could do so, we would be able to target circZNF609 by CRISPR/Cas9. We cloned the endogenous circZNF609 and recapitulated the expression of circZNF609 from its intron *in vitro*. However, we did not observe the circZNF609 generated from this construct (Supplemental Fig 5.10). Further analysis is necessary to understand how circZNF609 is generated from its endogenous locus.

### 3.6 Potential function of circZNF609 as miRNA sponges

Since a few circRNAs are well-known to function as miRNA sponges, circZNF609 were proposed to sponge 2 miRNAs: miR-615-5p and miR-150-5p in 2 different studies<sup>118,119</sup>. However, we would believe that such miRNA sponge situation may not happen widely in nature, not only for circZNF609 but also for many other circRNAs. This was suggested earlier by the fact that most circRNAs are low abundant<sup>69</sup>. Moreover, from the deep sequencing analysis, the expression of the miRNAs: miR-615-5p (Supplemental Fig. 5.6) and miR-150-5p in HT29 (Supplemental Fig. 5.7) is very low (less than 10 reads per million reads) and such low expression of miRNA may not be enough to function as miRNA sponges, at least in HT29 cells. We, therefore, believe that circZNF609 does not sponge the suggested miRNA in HT29 or at least such interaction would not be functionally relevant. We would speculate that if it happens, circZNF609 may sponge other miRNAs.

### 3.7 Analysis of proteins generated from circZNF609

Without any tag in overexpression constructs, we observed several protein products recognized by anti-Zfp609 antibodies apart from other nonspecific bands. In the stable cell line CMT93-circZfp609, we observed more than 2 protein products, some of which do not have corresponding protein bands in non-induction CMT93-circZfp609 cell line (Fig. 2.14 A). This could be explained by the fact that there are several proteins produced from the circRNA overexpression construct.

Indeed, we observed that several linear RNAs generated from circRNA overexpression constructs apart from the desired circRNAs. First, due to incorporation of strong bGH poly (A) stop signal at the end of the transcript, we could rule out the production of concatamer as discussed before (Fig. 2.7). However, there were at least 2 transcripts generated in this case: the linear precursor RNAs containing upstream intron-circularized exon-downstream intron and the circRNAs. Since the precursor RNAs contain the circularized exon with exact sequence of circZfp609, this could positively or negatively interfere with the function of circZfp609. For examples, in the overexpressed constructs, if we put the 3xFlag sequence after the first AUG of circZfp609, we observed 2 proteins products produced: one from the circRNA and other one from the linear precursor RNA (Fig. 2.12 D). In this case, we need to be very careful when analyzing the phenotype caused by any of these proteins or both.

### 3.8 Interaction partners of circZNF609 proteins

Identification the interaction partners of circZNF609 proteins gives us some hints about the function of circZNF609. Mass spectrometry analysis identified several proteins which belong to the integrator complex, including INTS10, INTS13, INTS14. Interestingly, integrator proteins were recently reported to interact with Zfp609 full-length protein to regulate gene expression<sup>115</sup>. Cellular fractionation revealed that the full-length Zfp609 localizes exclusively to the nucleus. Of note, the predicted NLS of Zfp609 is not present in the circularized exon of circZfp609. Consistently, we observed that the overexpressed linearized circZfp609 protein fused to Flag-HA localizes mainly to the cytoplasm.

If circZfp609 encodes for proteins, this cytoplasmic-localized proteins would interfere with the function of the full-length protein Zfp609. However, in this case, this also implies that Zfp609 protein may have function in colorectal cancer. This is not elucidated so far. Moreover, integrator complex proteins were reported to have additional function in the cytoplasm<sup>168</sup>. Therefore, the binding of circZfp609 to integrator complex proteins may function independently from Zfp609 proteins. Of note, integrator proteins are well-known for their function in snRNA biogenesis<sup>150,169</sup>. We checked the processing of snRNAs in case of circZNF609 knockdown but we have not observed any defect in biogenesis of snRNAs. Therefore, circZNF609 may have other functions than snRNA biogenesis. For examples, circZNF609 may sequester the integrator proteins to the cytoplasm and interfere other functions of integrator complex.

---

In case that circZNF609 does not encode for proteins and function as non-coding RNA, the interaction of the linearized part of circZNF609 and integrator proteins could also be explained. We would speculate that circZNF609 contains a protein part of full-length ZNF609, this part would serve as the domain for binding to integrator proteins. Overexpression of this truncated protein fragment from ZNF609 results in identification of integrator proteins from mass spectrometry data. This overexpression could give some "dominant-negative effect" to the cells because it localizes to the cytoplasm. Such hypothesis should be investigated further in the future. So far, the endogenous circZNF609 protein has not been found. Maybe circZNF609 protein are low abundant and therefore a relevant function of such low abundant protein is questionable. Further analysis using mass spectrometry analysis would be necessary to address this issue.

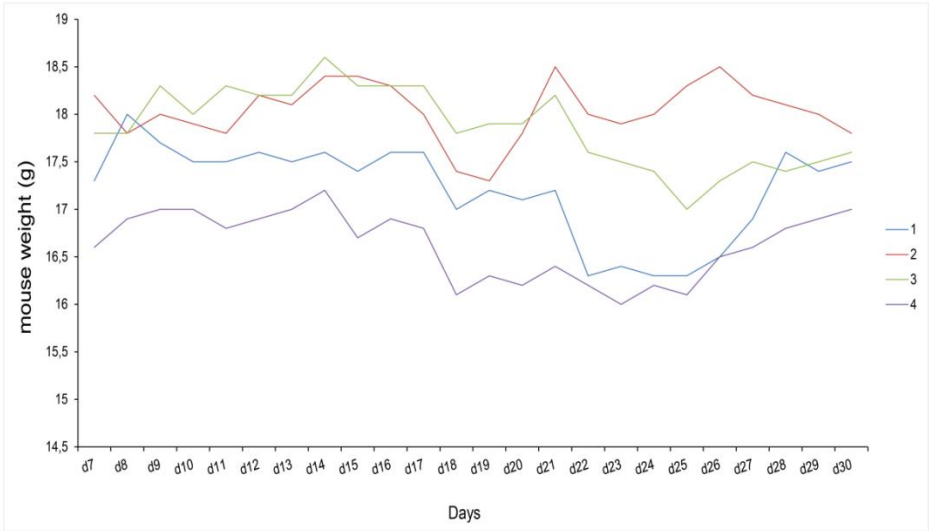
## 4 Outlook

There is still a long road ahead until the role of circZNF609 on colorectal cancer progression and metastasis becomes clear. Currently, we are performing cell proliferation, cell migration and cell invasion to understand the functions of circZNF609 and/or the proteins generated from circZNF609 *in vitro*. We are also performing xenograft mouse experiment of CMT93-Flag-HA-circZfp609 ORF transplanted into C57BL/6 to figure out the relevant function of circZNF609 *in vivo*. It is also interesting to understand if the full-length ZNF609 has any roles in colorectal cancer development. In this study, we found that both ZNF609 and circZNF609 proteins bind to some proteins of the integrator complex. In the future, it would be necessary to examine further this interaction in the context of colorectal cancer metastasis.

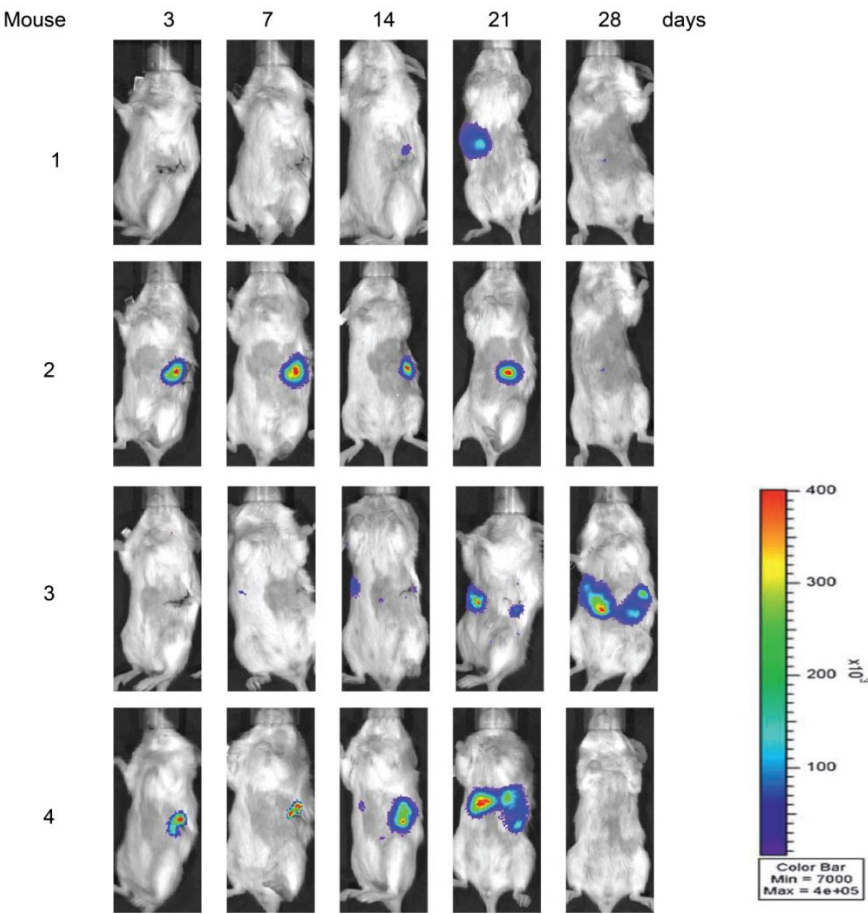
The mechanisms of how circRNAs can recruit ribosomes are not clear. At the moment, there are just only a few publications showing that circRNAs can encode for proteins<sup>84,86,87,88</sup>. Further experiments to identify initiation factors that recruit ribosomes for translation would help to understand circRNA translation mechanisms. In this work, we found that several AUGs can be used to initiate translation of circZfp609. It would be interesting to understand how AUGs can be selected for translation and how scanning mechanism works on circRNA in the future.

# 5 Appendix

A



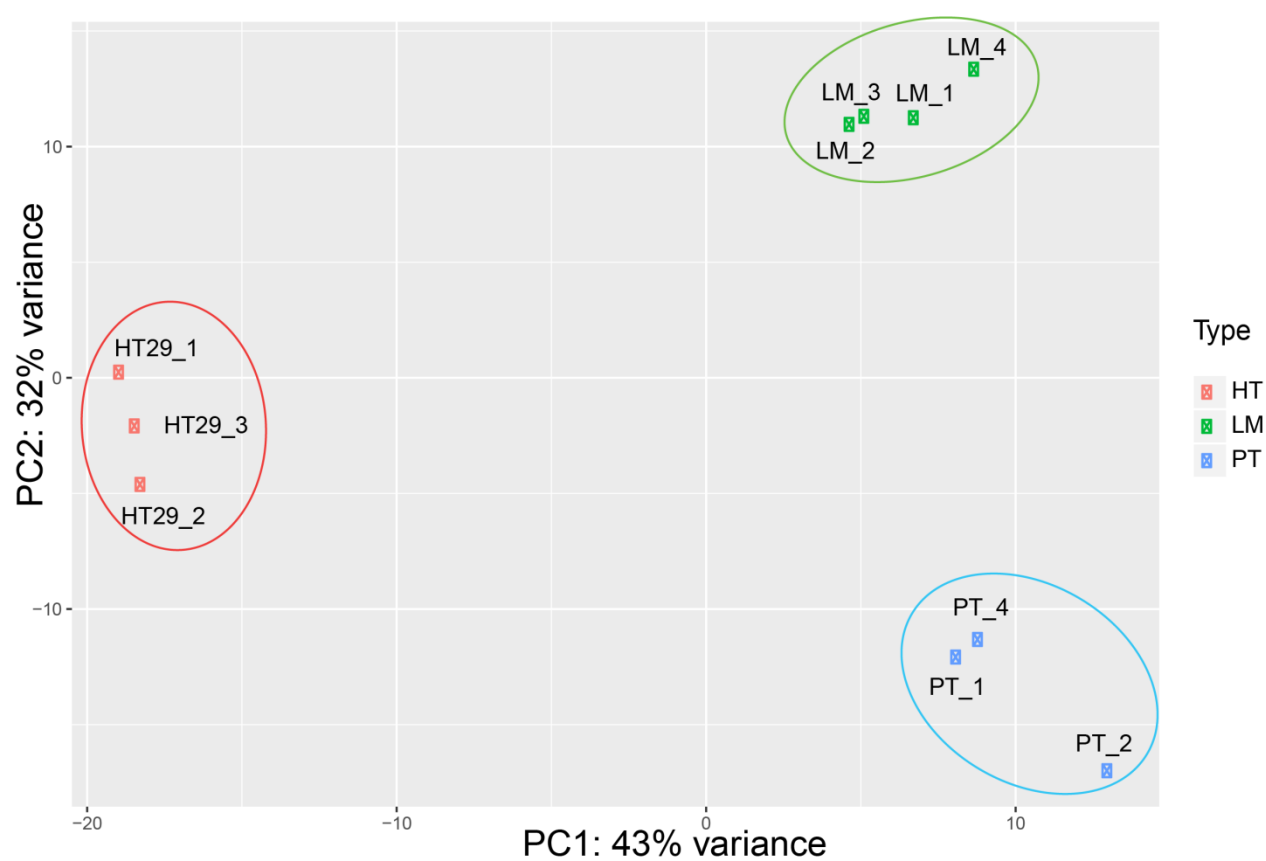
B



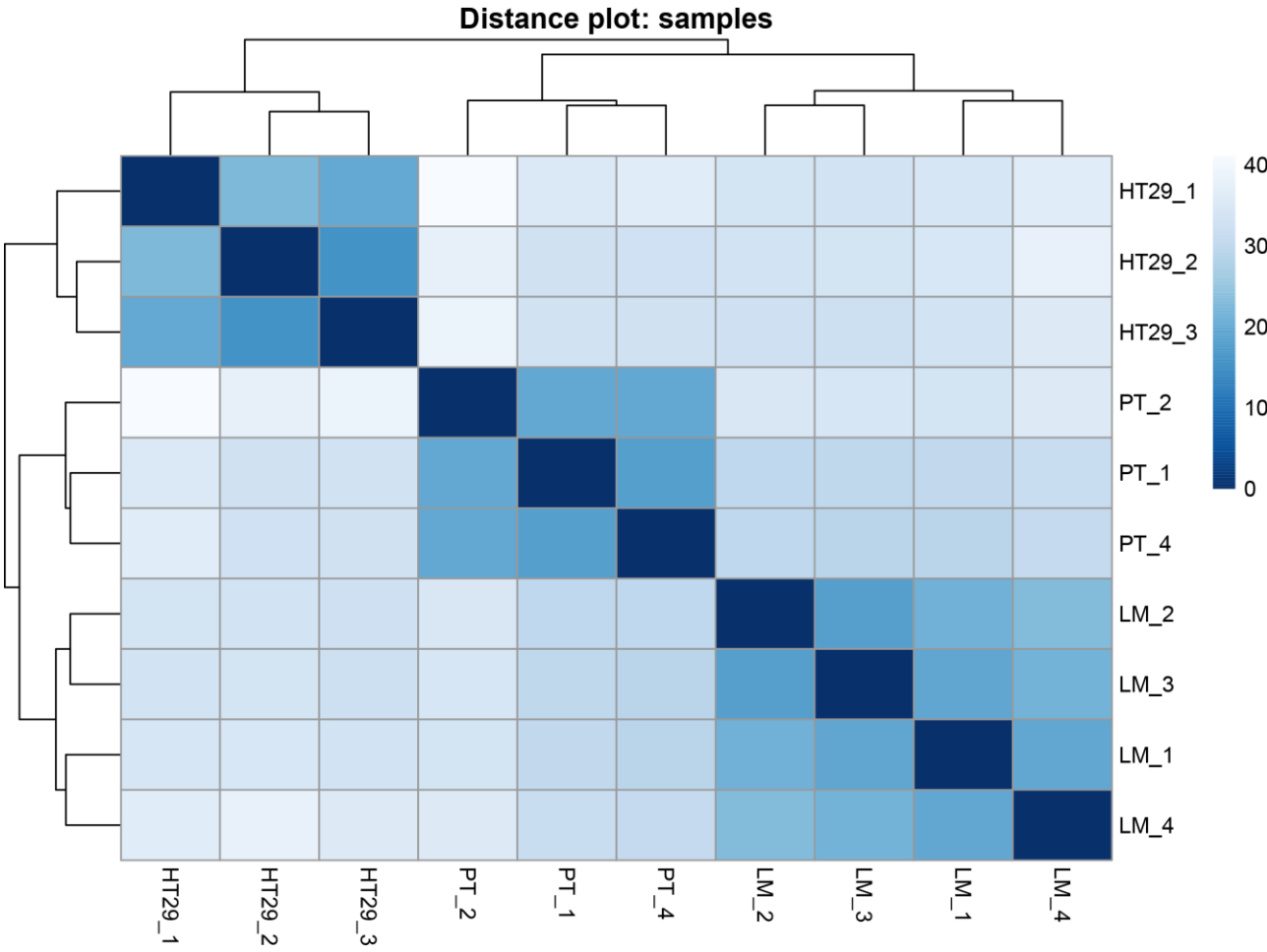
**Supplemental Fig. 5.1.** Mouse preparation for development of primary tumor and liver metastasis. A) Mouse weight (g) was measured every day 7 days after injection of HT29.hCG.Luc cells to the spleen. B) *In vivo* bioluminescence monitoring of intrasplenic HT29.hCG.Luc tumors. 3-7 days after injection, primary tumors developed. For bioluminescence imaging, mice were intraperitoneally injected with 150 mg/kg of luciferin and imaged weekly. All mice were sacrificed when bioluminescence levels reached about 500.000 photons/s (about 4 weeks after injection).

Sample	Tissue/Cell line	Cluster(Paired-Reads)	rRNA Read1	rRNA Read2	%rRNA	Mapped(hg19)
HT29-1	HT29	63.705.728	13.419.307	13.219.159	20,9%	56,5%
HT29-2	HT29	78.905.534	15.467.842	15.173.148	19,4%	52,0%
HT29-3	HT29	62.979.600	11.117.395	10.928.753	17,5%	52,9%
Mouse #1	PT	66.402.574	14.227.657	13.820.024	21,1%	69,9%
Mouse #2	PT	69.394.522	12.432.512	12.006.234	17,6%	62,1%
Mouse #4	PT	63.015.456	10.759.512	10.573.909	16,9%	74,0%
Mouse #1	LM	56.387.454	14.143.013	13.813.114	24,8%	71,3%
Mouse #2	LM	51.109.569	9.876.343	9.810.927	19,3%	75,3%
Mouse #3	LM	70.555.006	15.568.199	15.157.065	21,8%	69,6%
Mouse #4	LM	47.376.715	13.264.379	13.118.489	27,8%	73,3%

**Supplemental Fig. 5.2.** Information regarding to the deep sequencing of rRNA-depleted transcriptome library

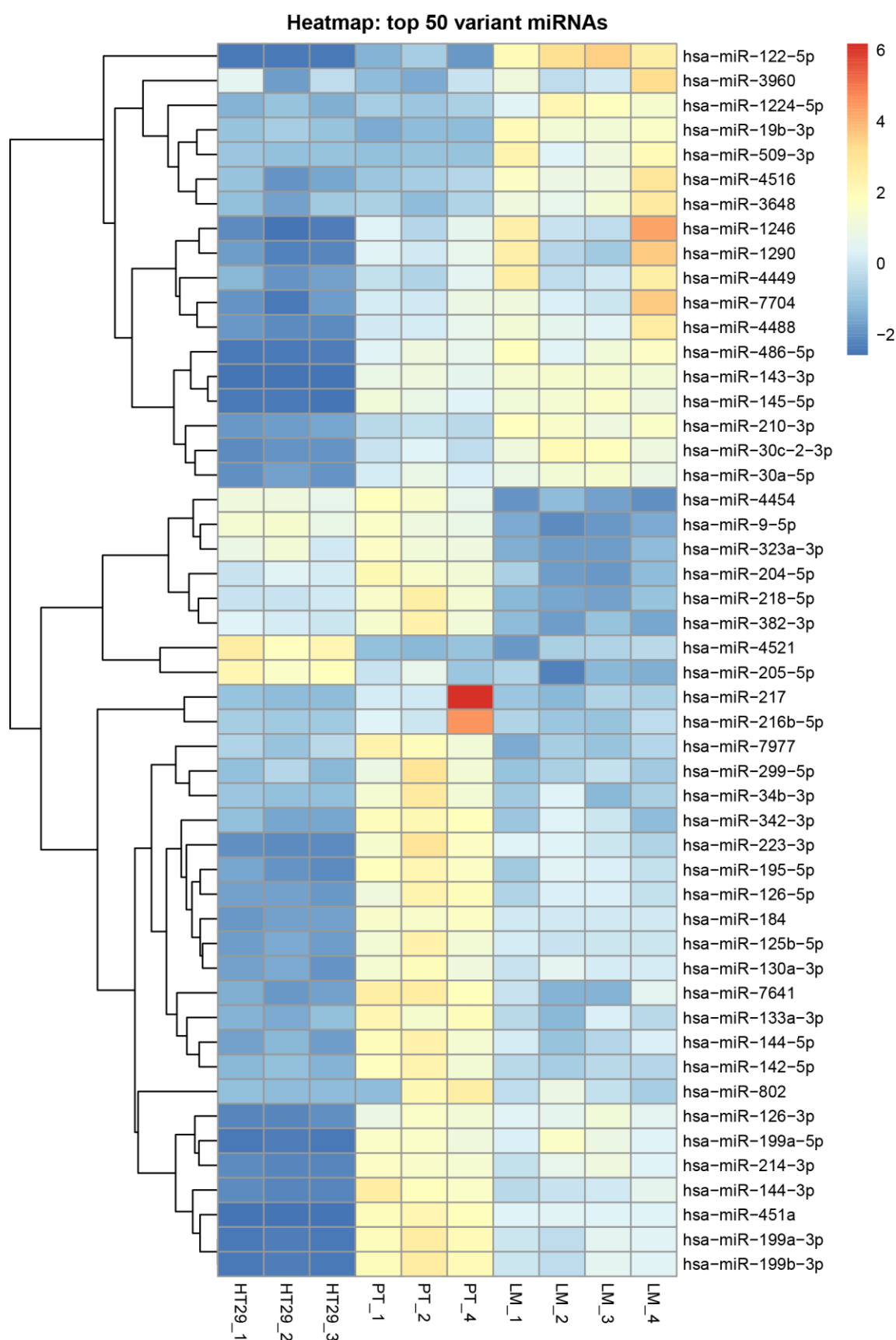


**Supplemental Fig. 5.3.** Principle component analysis (PCA) of all miRNA libraries prepared for deep sequencing from our mouse model. PT: primary tumors, LM: Liver metastasis.

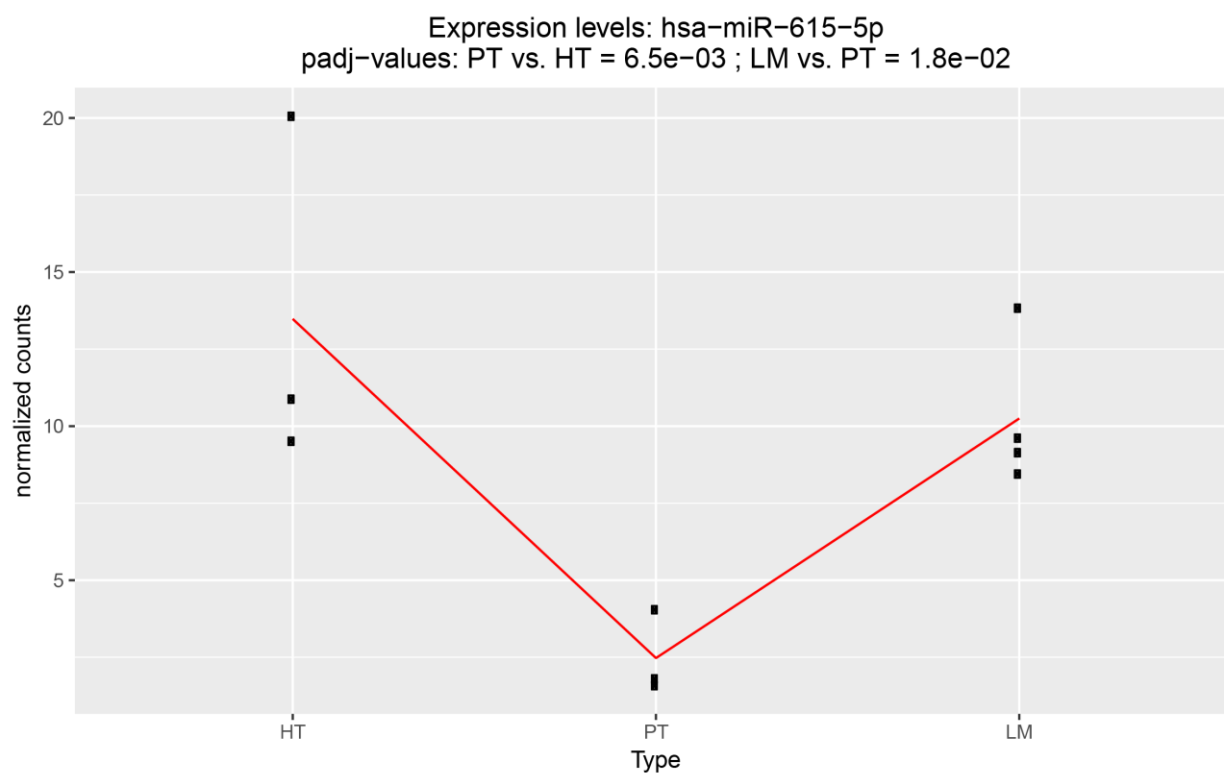


**Supplemental Fig. 5.4.** Distance plot of all miRNA libraries prepared for deep sequencing from our mouse model. PT: primary tumors, LM: Liver metastasis.

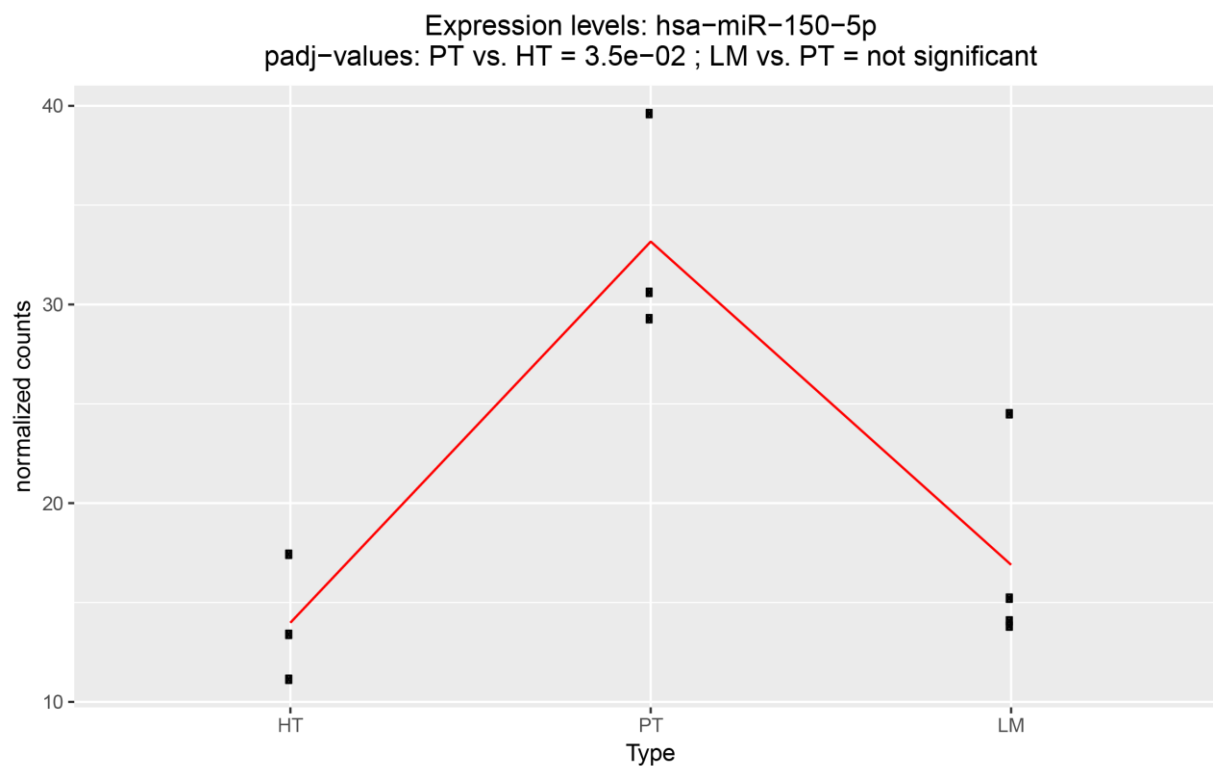




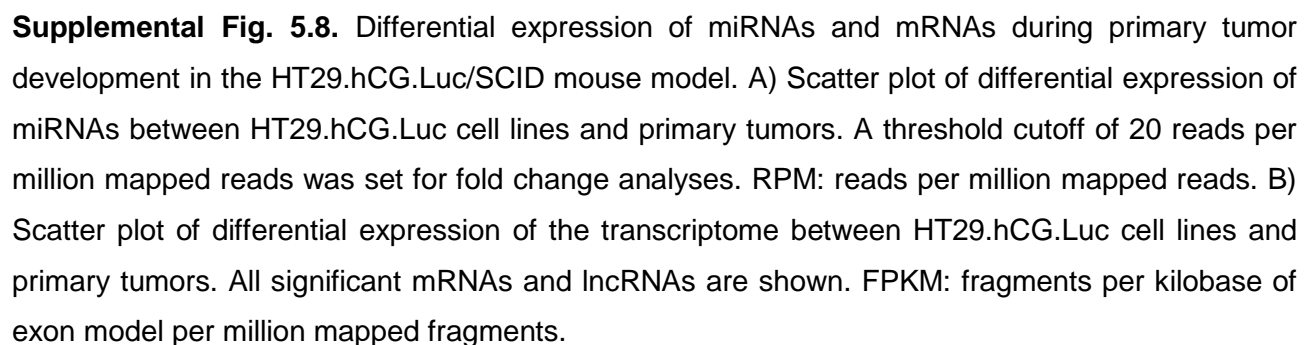
**Supplemental Fig. 5.5.** Heatmap of top 50 most variant miRNAs from deep sequencing data. PT: primary tumors, LM: Liver metastasis



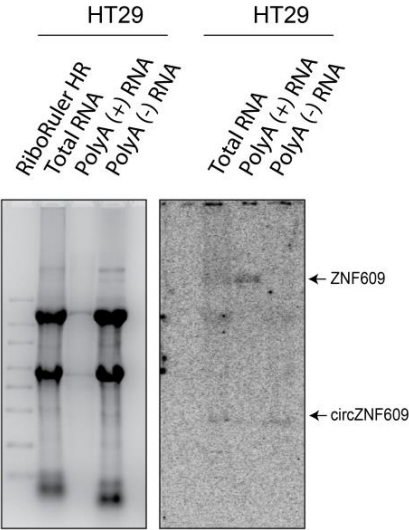
**Supplemental Fig. 5.6** Expression of miR-615-5p in HT29 cells (HT), primary tumors (PT) and liver metastasis (LM)



**Supplemental Fig. 5.7.** Expression of miR-150-5p in HT29 cells (HT), primary tumors (PT) and liver metastasis (LM)

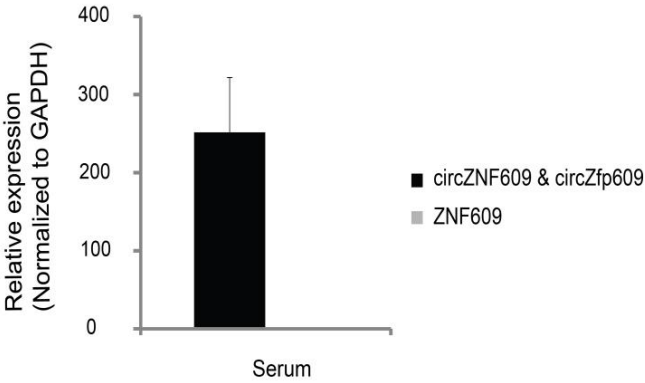


A



B

1% of human circZNF609 RNAs are secreted in to serum of SCID mice



C

circZNF609 1>GCAGCGCTCAATCCTTTGGGAAGTAAGCCGAGCCAGAGGAAGGGGAGAATGAGTG-CCGGCC-CTAAAGAAAGTCAAGTCTGAAAAGCAATGATGTTG>98

circZfp609 1>GCAGCGCTCAATCCTTTGGGAAGTAAGCCGAGCCAGAGGAAGGGGAGAATGAGTG-CCGGCC-CTAAAGAAAGTCAAGTCTGAAAAGCAATGATGTTG>98

99>TCCACTGGACATGTCTGACCAAGGAGGAGGCTCTGAGAACGTAGCTGAAGC>150

99>TCCACTGGACATGTCTGACCAAGGAGGAGGCTCTGAGAACGTAGCTGAAGC>150

D

HS MSLSSGASGGKGVNDANPVETDYGDEWDIGVGNLIIDLADLEKDQKLEMSGSKVEVGIP 60

MMU MSLSSGACGGKGVNDANPVETDYGDEWDIGVGNLIIDLADLEKDQKLEMSGSKVEVGIP 60

\*\*\*\*\*

HS APNAVATLPDNIKFVTPVPQPGQKEGKSKSKRSKSGKDTSKPTPGTSLFTPSEGAASKKE 120

MMU APNAVATLPDNIKFVTPVPQPGQKEGKSKSKRSKSGKDASKPTPGTSLFSPSEGAASKKE 120

\*\*\*\*\*

HS VQGRSGDGANAGGLVAAIAPKGSEKAAKASRSVAGSKKEKENSSSKSKKERSEGVGTCSE 180

MMU VQGRAGDGASAGGLVAAVAPKGSEKAAKASRSVAGSKKEKENSSSKGKKERSEGVGTCSE 180

\*\*\*\*\*

HS KDPGVLQPVPLGGRGGQYDGSAGVDTGAVEPLGSLAIEPGAALNPLGTPKPEPEEGENECR 240

MMU KDPGVLQPVPLGGRGSQYDGSAGMDTGTVEPLGSLAIEPGAALNPLGTPKPEPEEGENECR 240

\*\*\*\*\*

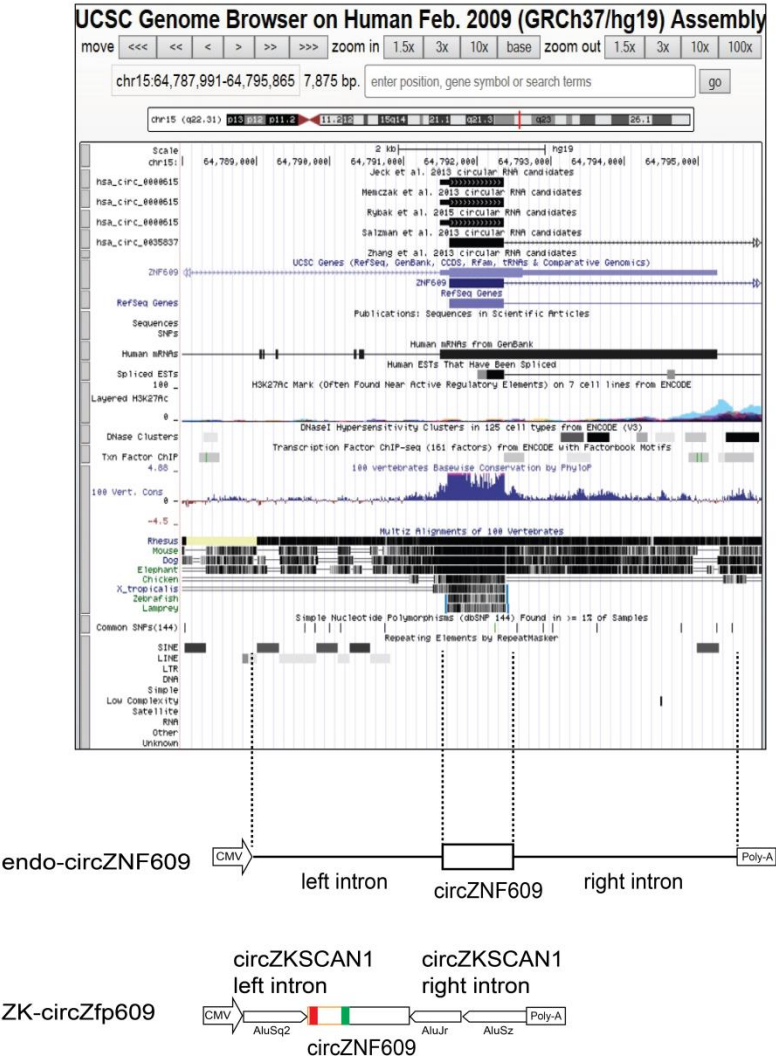
HS LLKKVKSEKQ 250

MMU PLKKVKSEKQ 250

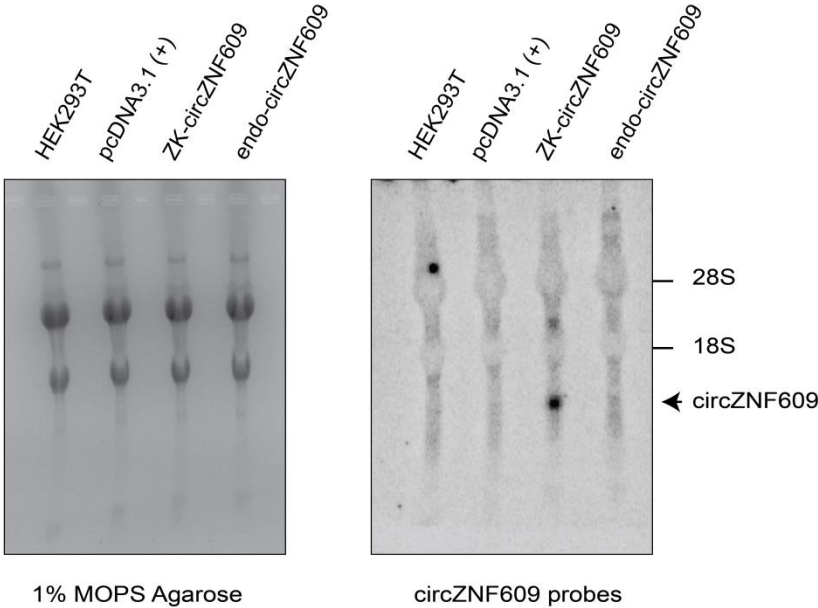
\*\*\*\*\*

**Supplemental Fig. 5.9. Characterization of circZNF609.** A) Northern blot of the fractionated polyA (+/-) RNA from HT29 cells. 20 ug of Total RNA or polyA (-) RNA and 2.5 ug of polyA (+) RNA was loaded on 1% MOPS Agarose gel. Northern blot using probes detected both ZNF609 and circZNF609 was performed. Northern blot analysis showed that circZNF609 is not enriched in the polyA (+) fraction as shown by ZNF609. We detected circZNF609 in both Total RNA or polyA(-) RNA fraction. B) Serum of the SCID mice injected with HCT116.hCG.Luc cells was collected at the time of scarification (around 4 weeks after injection). Relative expression using qPCR primer for circZNF609 showed the abundance of its expression in serum normalized to GAPDH. Notably, linear RNAs including GAPDH and ZNF609 are not stable in the serum. Since circZNF609 is highly conserved, the qPCR could presumably detect the murine circZfp609. C) Alignment of the sequence of the amplicon of circZNF609 and circZfp609 that amplified by qPCR primer. Since there are mismatches between human circZNF609 and murine circZfp609, deep sequencing showed that 0,84 % of the amplicon belongs to human circZNF609. This implied that the human circZNF609 is secreted to the serum of the SCID mice and potentially served as future promising biomarkers. D) Alignment of the aa sequence of human circZNF609 (HS) and murine circZfp609 (MMU).

A

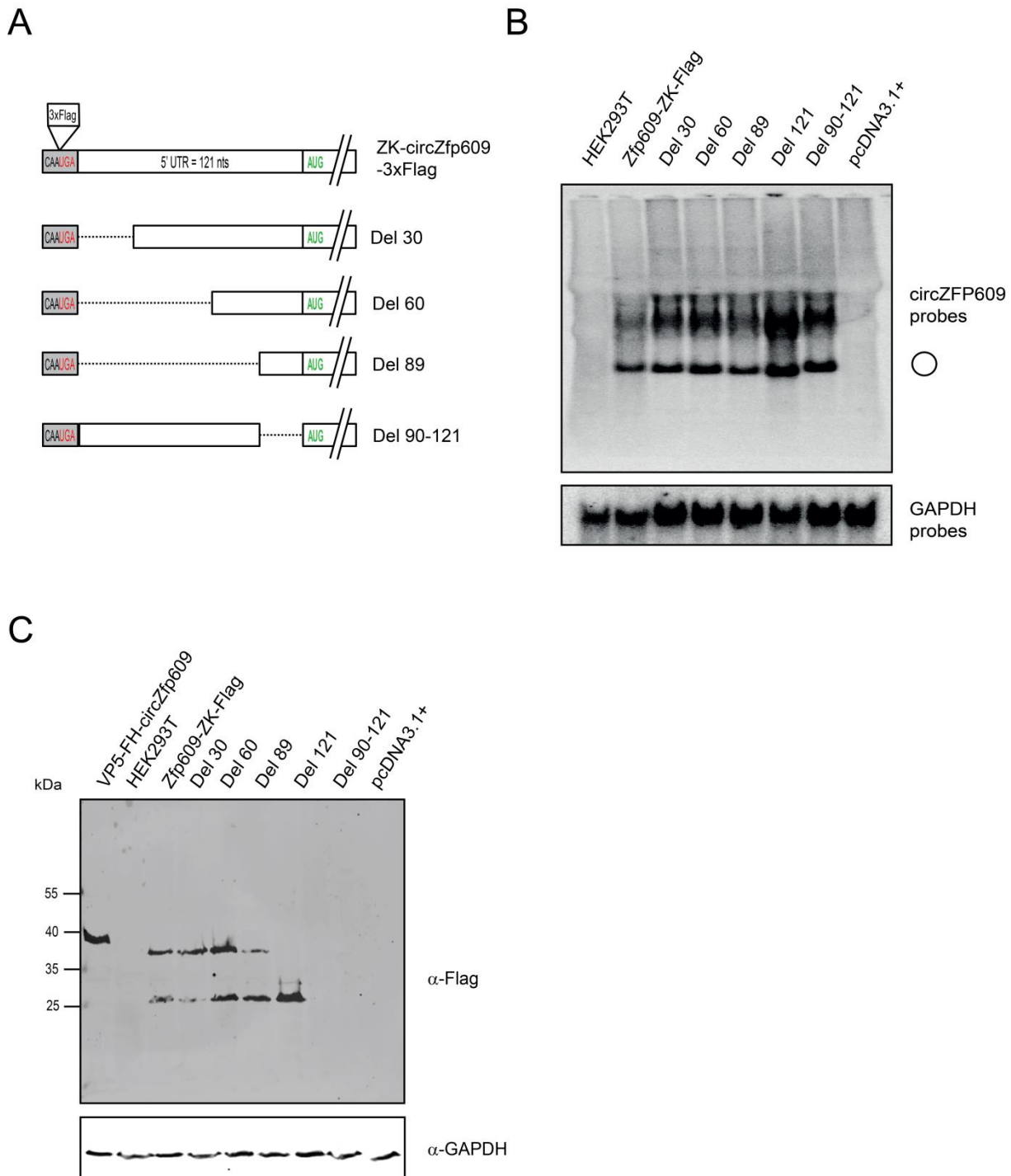


B



**Supplemental Fig. 5.10. Analysis of circZNF609 produced from its endogenous locus.** A) Snapshot of the UCSC genome browser of the circZNF609 locus. Here, it was annotated that circZNF609 is generated from exon 1 of human ZNF609. However, from our RNAseq analysis, circZNF609 comes from exon 2 of ZNF609. circZNF609 was previously identified from several studies<sup>27, 26, 65</sup>. PCR amplifying 4 kb upstream intron, circZNF609 exon and 3 kb downstream intron was performed. Endo-circZNF609 was generated by assembly these fragments by SLIC cloning. The full fragment was then cloned into the pcDNA3.1 (+) vector. B) HEK293T was transfected with different constructs for expressing circZNF609. 2 days post-transfection, the RNA was harvested for northern blot analysis. Northern blot showed the strong expression of the positive control, circZNF609 expression form ZK-circZNF609 vector. However, the endo-circZNF609 showed no signal of circZNF609.

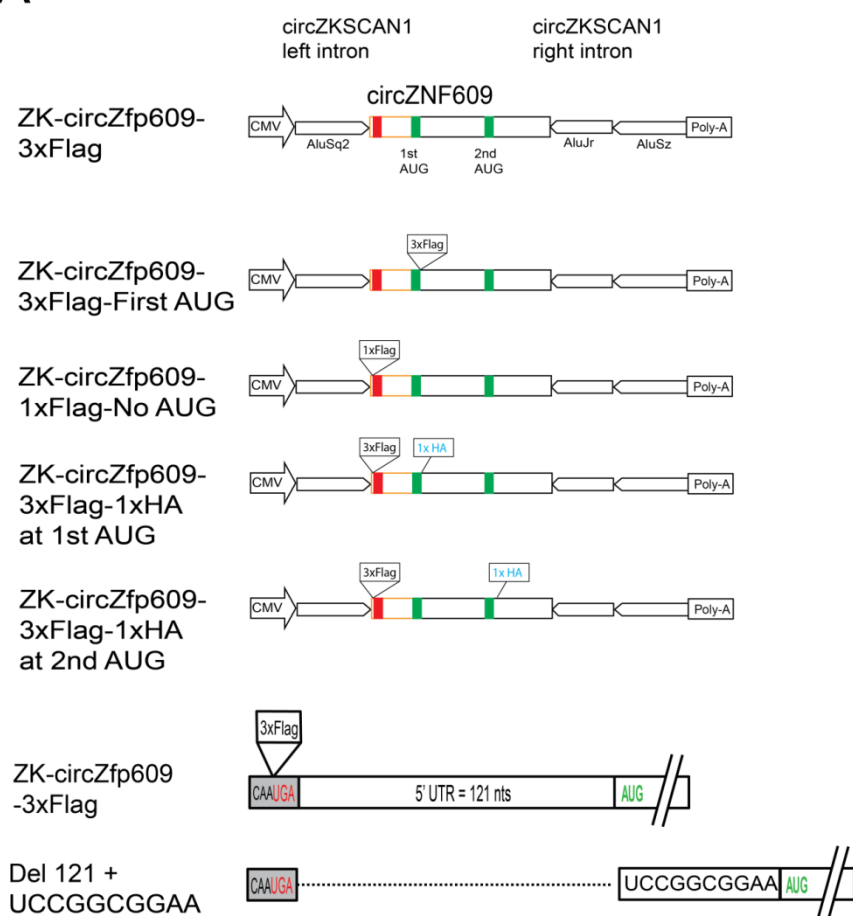




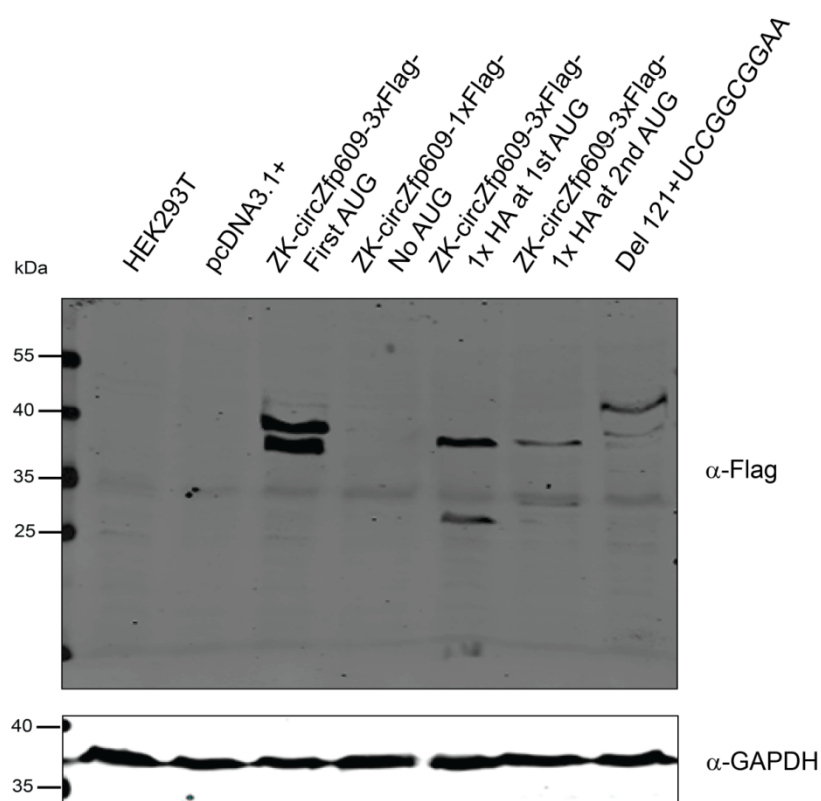
**Supplemental Fig. 5.11.** Further mutagenesis of Zfp609 5' UTR for studying its translation mechanism. A) Truncated mutants of circZfp609 5' UTR. C) Northern blot of HEK293T cells transfected with all 5' UTR deletion mutants showed expression of overexpressed circZfp609 5' UTR mutants. D) Western blot of samples from C) with anti-Flag antibodies. Several mutants were shown in Fig. 2.9. Only Del 90-121 was included in this western blot. The data show that Del 90-121 gives us no protein expression recognized by anti-Flag antibodies.



A

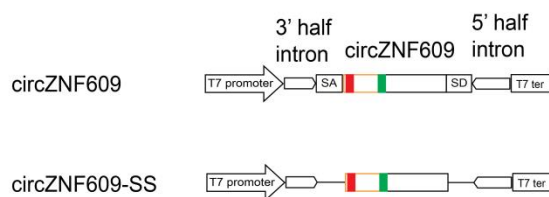


B

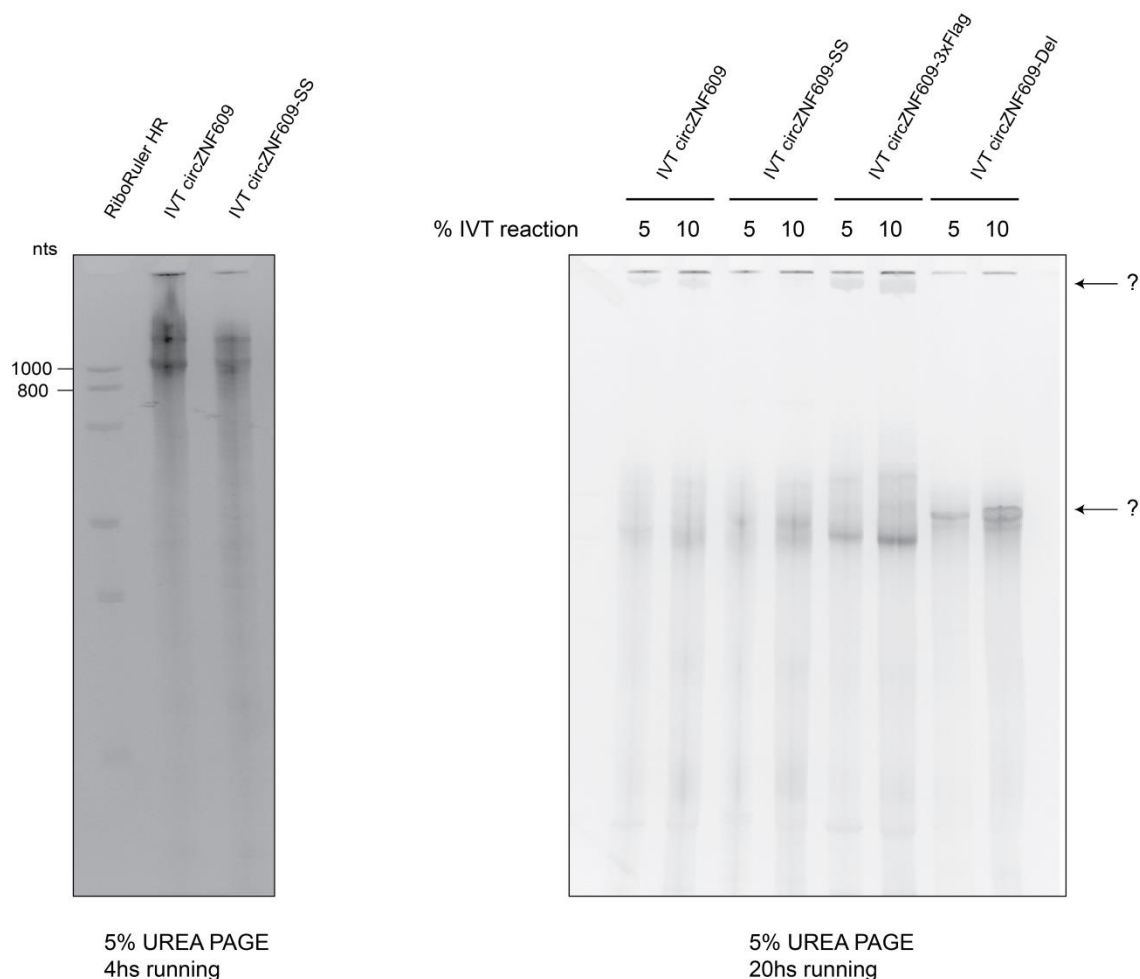


**Supplemental Fig. 5.12.** Further mutagenesis of Zfp609 UTR for studying its translation mechanism (continued). A) Truncated mutants of circZfp609 5' UTR were generated to understand how circZfp609 can recruit the ribosome. 1x Flag without any AUG in its sequence was inserted just before the stop codon (ZK-circZfp609-1xFlag No AUG). 1x HA was inserted into the circ-Zfp609-3x Flag just after the first AUG (ZK-circZfp609-3xFlag-1xHA at 1st AUG) or second AUG (ZK-circZfp609-3xFlag-1xHA at 2<sup>nd</sup> AUG). 10 unrelated nts were added to the Del 121 construct. B) Western blot of the HEK293T cells transfected with different mutants by anti-Flag antibodies. ZK-circZfp609-1xFlag No AUG gave no protein band. This implied that 3xFlag is better for recognition by anti-Flag Antibodies. ZK-circZfp609-3xFlag-1xHA at 1<sup>st</sup> AUG give 2 protein products. Surprisingly, ZK-circZfp609-3xFlag-1xHA at 2<sup>nd</sup> AUG gave only one bigger protein products. The second protein products were not detected by anti-Flag antibodies. As discussed, potential AUG in HA was removed but unexpectedly, one stop codon in the HA sequence could potentially use from upstream AUG resulting one complete ORF. We speculated that this ORF was used and therefore the second AUG of Zfp609 was not used for translation. Further analysis should be performed to get a final conclusion about these mutants. Del 121 gave lower protein products from second AUG. However, Del 121+UCCGGCGGAA gave similar protein products as Del 121-AU.

A

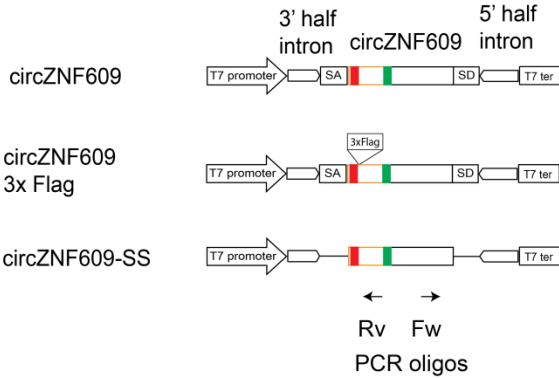


B

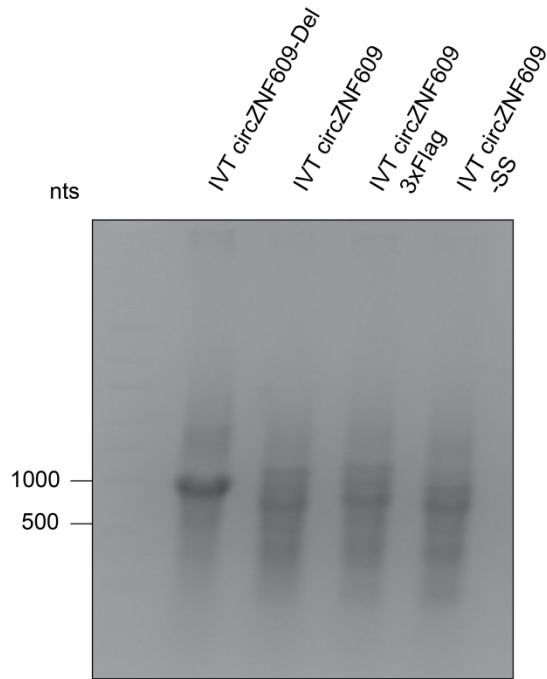


**Supplemental Fig. 5.13.** IVT for synthetic circZNF609 production on Urea PAGE. A) Scheme of constructs using the PIE methods for production of circZNF609. circZNF609-SS was generated by deleting 9 splice sites necessary for *in vitro* self-splicing reaction by quick change mutagenesis. The RNA was isolated after IVT reaction using TRIzol-LS reaction. RNA was run on 6% Urea PAGE for 4 or 20 hours.

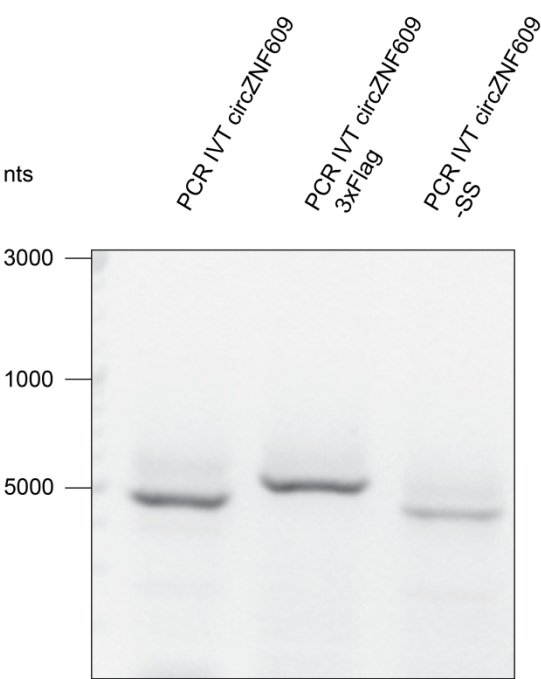
A



B



C



D

Sequence alignment of circZNF609 ref and circZNF609-SS. The alignment shows the reference sequence (circZNF609 ref) and the modified sequence (circZNF609-SS) with asterisks indicating the positions of the modifications.

circZNF609 ref: 1>TAAACCGGAGCCAGAGGAAGGGGAGAATGAGTGTGCGCTGCTAAAGAAAGTCAAGTCTGAAAAGCAATGATGTTGTCCACTGGGCATGTACTGACCAATG>100

circZNF609-SS: 1>TAAACCGGAGCCAGAGGAAGGGGAGAATGAGTGTGCGCTGCTAAAGAAAGTCAAGTCTGAAAAGCAATGATGTTGTCCACTGGGCATGTACTGACCAATG>92

101>TGGCAGGCTCTGAGAACATAGCTGAAGCTGAAAATAGGAAAGCTGGGGCAAGGAAGAGCCTTGAATCTTGAGGTGGGACGTTGACTCTAAGATGTCTTG>200

93>TGGCAGGCTCTGAGAACATAGCTGAAGCTGAAAATAGGAAAGCTGGGGCAAGGAAGAGCCTTGAATCTTGAGGTGGGACGTTGACTCTAAGATGTCTTG>192

201>AGCAGTGGAGCCTCCGGAGGGAAAGGAGTGGATGCAAAACCGGTTGAGACATACGACAGTGGGGATGAATGGGACATTGGAGTAGGGAATCTCATCATTG>300

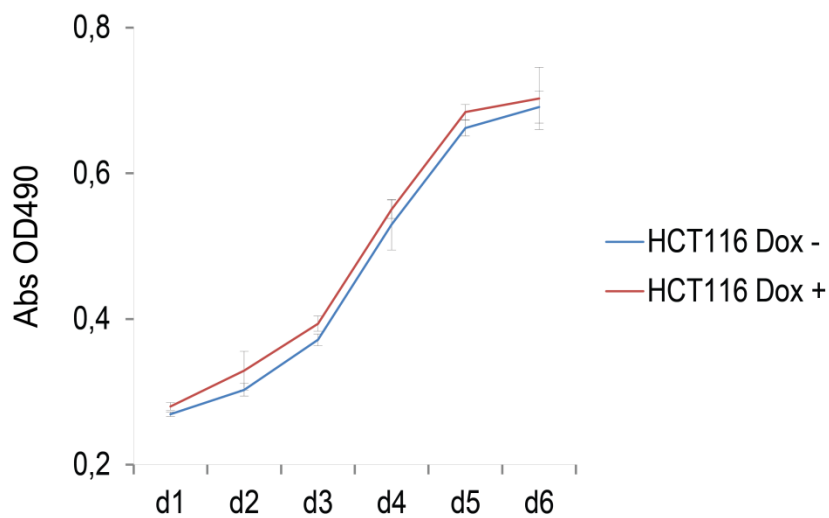
193>AGCAGTGGAGCCTCCGGAGGGAAAGGAGTGGATGCAAAACCGGTTGAGACATACGACAGTGGGGATGAATGGGACATTGGAGTAGGGAATCTCATCATTG>292

301>ACCTGGACGCCGATCTGAAAAGGACCGAGCAAACTGGAATGTGAGGCTCAAAGGAGGTGGGGATACCGGCTCCCAATGCTGTGGCCACACTACCAAG>400

293>ACCTGGACGCCGATCTGAAAAGGACCGAGCAAACTGGAATGTGAGGCTCAAAGGAGGTGGGGATACCGGCTCCCAATGCTGTGGCCACACTACCAAG>392

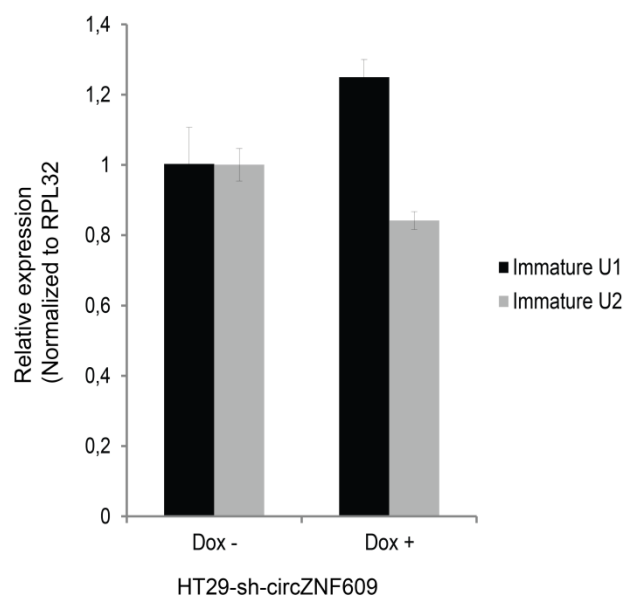
**Supplemental Fig. 5.14.** Fidelity of the synthetic circZNF609 by PIE method. A) Scheme of PCR oligos for amplification of synthetic circZNF609. B) The RNA was isolated after IVT and ran on 1% MOPS Agarose. The data showed that circZNF609-SS RNAs are generated from self-splicing using PIE method. C) 1 ng of the isolated RNA from MOPS Agarose gel was used for cDNA synthesis. PCR of the indicated oligos was performed. The PCR products were run on 1% Agarose gel. D) Isolated PCR products were then cloned into a pGEMT-easy vector. Sanger sequencing showed that the RNA obtained from circZNF609-SS lacks 8 nts of the circZNF609 junction.

**A**



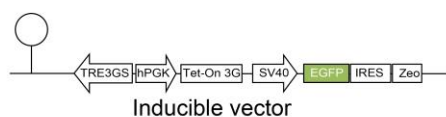
**Supplemental Fig. 5.15.** Doxycycline does not have any effect on the proliferation of HCT116 cells.

A

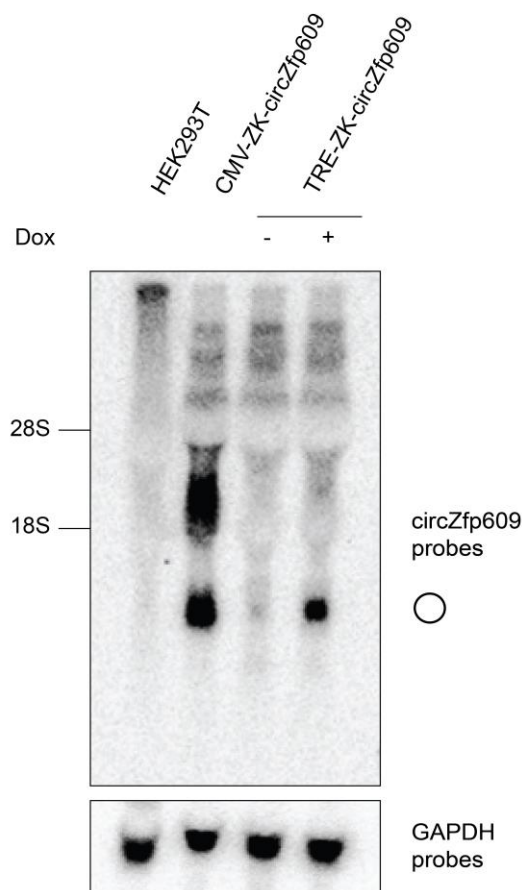


**Supplemental Fig. 5.16.** Knocking down of circZNF609 does not have effect in the processing of U1 and U2 snRNAs. HT29-sh-circZNF609 cells were induced with doxycycline for 3 days. After that, RNA was collected for northern blot and qPCR analysis. Northern blot analysis showed no changes in the expression of premature U1, U2 and U4 snRNAs (data not shown). qPCR was then used to quantitatively calculate the changes in the immature U1 and U2 snRNAs. Comparing to the non-induced HT29-sh-circZNF609 cells, knocking down of circZNF609 gave very small effect to the expression of premature U1 and U2 snRNAs.

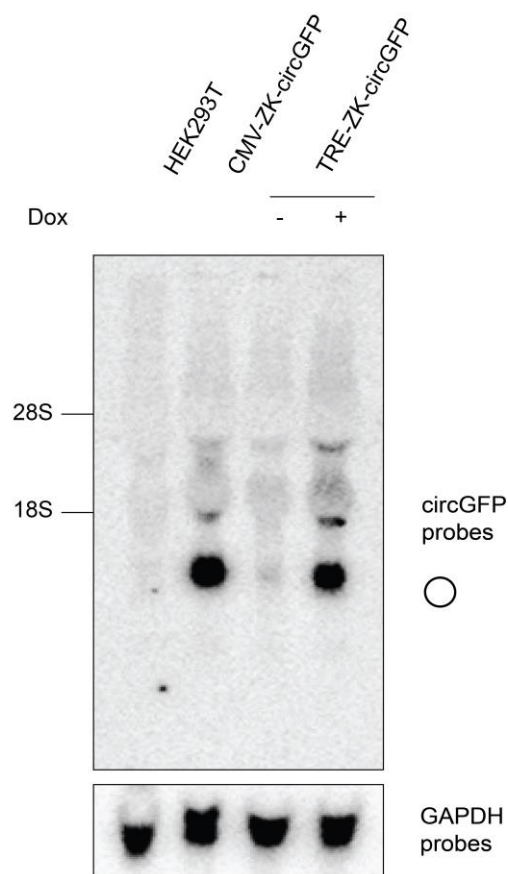
A



B



C



**Supplemental Fig. 5.17.** Method for inducible overexpression of circRNA. A) Scheme of vector for inducible overexpression of circRNAs. B) HEK293T was transfected with vector consecutively expressing circZfp609 (CMV-ZK-circZfp609) or vector for inducible expression of circZfp609 (TRE-ZK-circZfp609). 2 days post-transfection, RNA was isolated for northern blot analysis. Northern blot showed expected expression of circZfp609 upon doxycycline induction. C) The same experiment was performed with the vector for overexpression of circGFP.

**Table 1. List of abbreviations**

aa	amino acid
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
Ago	Argonaute
Amp	ampicillin
APS	ammonium persulphate
ATP	adenosine triphosphate
bp	base pair(s)
BSA	bovine serum albumin
ceRNA	competing endogenous RNA
cDNA	complementary DNA
CoIP	co-immunoprecipitation
circRNA	circular RNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DANN	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
Dox	Doxycycline
ds	double-stranded
DTT	dithiothreitol
EDC	1-ethyl-3-(3-dimethyl-aminopropyl)-carbodiimid
EDTA	ethylenediaminetetraacetic acid
eIF	eukaryotic initiation factor
EtBr	ethidium bromide
Fig.	Figure
FBS	fetal bovie serum
FPKM	fragments per kilobase of exon model per million
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GDP	guanosine diphosphate
GTP	guanosine triphosphate
hCG	Human chorionic gonadotropin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid



---

HygB	Hygromycin B
IPTG	isopropyl $\beta$ -D-1-thiogalactopyranoside
IRES	internal ribosome entry site
IVT	<i>in vitro</i> transcription
kDa	kilodalton
LM	Liver metastasis
Luc	Luciferase
m6A	N6-methyladenosine
M	molar
MCS	multiple cloning site
min	minute
miRNA	microRNA
RNP	ribonucleoprotein
mRNA	messenger RNA
NLS	nuclear export signal
nt (s)	nucleotides(s)
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PIE	permuted intron-exon
pri-miRNA	primary miRNA
PT	Primary tumor
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNP	ribonucleoprotein
RPM	reads per million
RT	room temperature
SA	splice acceptor
SCID	severe combined immunodeficiency
SD	splice donor
sec	second
shRNA	short hairpin RNA
siRNA	small interfering RNA

---

SDS	sodium dodecyl sulfate
SSC	saline-sodium citrate bu.er
TBE	Tris/Borate/EDTA buffer
TBS	Tris buffered saline
TBS(-T)	Tris-buffered saline (containing Tween 20)
TEMED	tetramethylethylenediamine
UTP	uridine triphosphate
UTR	untranslated region
XTT	2,3-Bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2H-Tetrazolium-5-Carboxanilide

**Table 2. circRNAs were reported to function as miRNA sponges.**

No.	Name of the circRNA	miRNA sponge	Target	Tissue relevant or Diseasea	Refs
1	HRCR	miR-223	ARC	heart	170
2	CircRNA-CER	MiR-136	MMP13	osteoarthritis	171
3	CircHIPK3	miR-124	IL6R and DLX2	Human cell growth	62
4	CircRNA_001569	miR-145	E2F5, BAG4 and FMNL2	Colorectal cancer	172
5	Hsa_circ_0005075	miR-23b-5p, miR-93-3p, miR-581, miR-23a-5p	Several	HCC	173
6	CircRNA MYLK	miRNA-29a-3p	DNMT3B, VEGFA and ITGB1	bladder	174
7	CircTCF25	miR-103a-3p/miR-107	CDK6	bladder	175
8	CircRar1	miR-671	Caspase8 and p38	Express in hippocampus and cerebral cortex Neurotoxicity	176
9	CircZNF609	miR-150-5p	AKT3	Hirschsprung disease	118
10	CircPVT1	miR-125 family		Gastric cancer	177
11	Circ100284	miR-217	EZH2	arsenite-induced cell cycle, express in keratinocytes	178
12	CircRNA_000203	miR-26b-5p	Col1a2, CTGF	mouse cardiac fibroblasts	179
13	CircHIAT1	miR-195-5p/29a-3p/29c-3p	CDC42	clear cell renal cell carcinoma	180

14	CircCCDC66	miR-33b, miR-93	MYC	Colon Cancer	181
15	Hsa_circ_0005105	miR-26a	NAMPT	Osteoarthritis	182
16	CircRNA_100290	miR-29b	CDK6	oral cancer	183
17	CircRNA_010567	miR-141	TGF- $\beta$ 1	myocardial fibrosis	184
18	MFACR	miR-652-3p	MTP18	cardiomyocyte death	
19	CircMTO1	microRNA-9	p21	HCC	185
20	Hsa_circ_0010729	miR-186	HIF-1 $\alpha$	vascular endothelial cell	186
21	CircRNA_100269	miR-630	no	gastric cancer	187
22	MYLK	miR-29a	VEGFA/VEGFR2	bladder	188
23	CircRNA_000839.	miR-200b	RhoA	hepatocellular carcinoma	189
24	Hsa_circ_0020397	miR-138	TERT and PD-L1	colorectal cancer	190
25	CircRNA_100338	miR-141-3p	no		191
26	CircABCB10	miR-1271	no	Breast cancer	192
27	Hsa_circ_0045714	miR-193b	IGF1R	osteoarthritis	193
28	CircLARP4	miR-424-5p	LATS1	Gastric cancer	194
29	hsa_circ_0001982	miR-143	no	Breast Cancer	195
30	9 circRNAs	miR-130a-3p	TRPM3	coronary artery disease	196
31	UBAP2	miR-143	Bcl-2	osteosarcoma	197
32	CircRNA_0046367	miR-34a	PPAR $\alpha$	Hepatic steatosis	198

33	CircGFRA1	miR-34a	GFRA1	breast cancer	199
34	Mmu_circRNA_007893	mmu-miR-485-5p	IL-6	macrophages	200
35	CircWDR77	miR-124	FGF-2	vascular smooth muscle cells	201
36	CircFUT10	miR-133a	MyoD, MyoG and MyhC	myoblasts	202
37	CircLMO7	miR-378a-3p	HDAC4	myoblasts differentiation	203
38	CircBIRC6	miR-34a, and miR-145	several	undifferentiated human embryonic stem cells	204

**Table 3. Potential human targets of circZNF609 siRNA (from <https://blast.ncbi.nlm.nih.gov/Blast.cgi>)**

Potential human targets	Potential mouse targets
<ul style="list-style-type: none"> <li>• UBE2V1</li> <li>• ZNF609</li> <li>• GLB1L2</li> <li>• CAMTA1</li> <li>• FNAR1</li> <li>• LOC105370456</li> <li>• FARSB</li> </ul>	<ul style="list-style-type: none"> <li>• Zfp609</li> <li>• Lvrn</li> <li>• Slc16a1</li> <li>• GM30771</li> <li>• Phf21a</li> <li>• Dmd</li> <li>• Hus1</li> <li>• Cdc42</li> <li>• Hipk1</li> </ul>

**Table 4. Top 50 proteins identified from Mass spec analysis of Flag IP from HEK293T-circZNF609 ORF upon doxycycline induction.**

No.	Protein name	Accession	Scores	#Peptides	SC [%]
1	INTS13	sp Q9NVM9 ASUN_HUMAN	5071.7	76	70,3
2	INTS10	sp Q96SY0 VWA9_HUMAN	3677.9	46	77,4
3	ZNF609	tr A5PKW6 A5PKW6_HUMAN	3426.7	39	61,9
4	INTS14	tr H3BRY6 H3BRY6_HUMAN	3087.5	39	72,2
5	HEL-S-89	tr V9HWP4 V9HWP4_HUMAN	3059.3	40	42,5
6	PPM1B	sp O75688 PPM1B_HUMAN	2903.3	35	67,8
7	KIF11	sp P52732 KIF11_HUMAN	2731.6	48	45,2
8	TUBB	tr Q5SU16 Q5SU16_HUMAN	2528.9	40	65,8
9	TUBB4B	sp P68371 TBB4B_HUMAN	2428.7	40	65,6
10	STK38	tr A0A024RD18 A0A024RD18_HUMAN	2340.7	36	57,2
11	STK38L	sp Q9Y2H1 ST38L_HUMAN	2127.5	31	58,2
12	WDR77	tr A0A024R0H7 A0A024R0H7_HUMAN	2119.3	29	68,1
13	TUBB4A	sp P04350 TBB4A_HUMAN	2100.6	35	61,9
14	INTS10	sp Q9NVR2 INT10_HUMAN	1935.5	37	45,6
15	KCTD5	sp Q9NXV2 KCTD5_HUMAN	1818.3	20	69,7
16	N/A	tr B2RBD5 B2RBD5_HUMAN	1676.2	26	29,8
17	C11orf84	sp Q9BUA3 CK084_HUMAN	1479.5	19	66,9
18	N/A	tr B3KPS3 B3KPS3_HUMAN	1428.7	21	57,2
19	KCTD17	sp Q8N5Z5 KCD17_HUMAN	1361.1	19	52,6
20	HEL-S-72p	tr V9HW22 V9HW22_HUMAN	1330.8	28	37,2
21	MAP3K7IP1	tr A8K6K3 A8K6K3_HUMAN	1313.1	25	65,3
22	KCTD2	sp Q14681 KCTD2_HUMAN	1288.8	15	48,7
23	UNC84B	tr A0A024R1P7 A0A024R1P7_HUMAN	1287.3	23	48,5
24	TUBA1C	sp Q9BQE3 TBA1C_HUMAN	1276.6	20	54,3
25	C7orf26	sp Q96N11 CG026_HUMAN	1035.4	12	31,6

26	RIOK1	sp Q9BRS2 RIOK1_HUMAN	986.0	21	28,5
27	FLJ51907	tr B7Z4V2 B7Z4V2_HUMAN	985.5	19	29,0
28	N/A	tr A0A140VJQ4 A0A140VJQ4_HUMAN	920.2	16	44,2
29	PABPC1	tr A0A024R9C1 A0A024R9C1_HUMAN	892.2	19	32,5
30	EIF4B	sp P23588 IF4B_HUMAN	803.6	12	31,9
31	CLNS1A	sp P54105 ICLN_HUMAN	784.7	12	51,1
32	PRMT5	sp O14744 ANM5_HUMAN	777.1	16	25,3
33	DHX9	sp Q08211 DHX9_HUMAN	766.1	18	17,7
34	TRIM21	sp P19474 RO52_HUMAN	759.1	16	32,2
35	FLJ35376	tr B3KS36 B3KS36_HUMAN	753.1	13	25,4
36	N/A	tr Q5U0D0 Q5U0D0_HUMAN	751.6	18	31,6
37	INTS13	tr F8VRX9 F8VRX9_HUMAN	745.4	11	73,4
38	PRPF31	tr F1T0A5 F1T0A5_HUMAN	718.8	13	34,7
39	PRMT1	tr E9PKG1 E9PKG1_HUMAN	686.6	16	51,4
40	TAB3	sp Q8N5C8 TAB3_HUMAN	677.4	13	28,4
41	FLJ53619	tr B4DGL0 B4DGL0_HUMAN	655.2	15	24,5
42	HNRNPM	sp P52272 HNRPM_HUMAN	651.9	13	17,7
43	SPIN	tr A0A024R297 A0A024R297_HUMAN	648.6	12	50,0
44	RPL7A	sp P62424 RL7A_HUMAN	620.4	12	26,3
45	HNRNPH1	sp P31943 HNRH1_HUMAN	610.1	10	33,4
46	PABPC4	tr B1ANR0 B1ANR0_HUMAN	594.8	11	19,5
47	FLJ55253	tr B4DW52 B4DW52_HUMAN	586.0	11	36,6
48	TCP1	sp P17987 TCPA_HUMAN	584.9	16	34,0
49	HEL-S-97n	tr V9HW63 V9HW63_HUMAN	581.6	11	45,4
50	N/A	tr A8K4Z4 A8K4Z4_HUMAN	571.8	7	36,0
51	FLJ55635	tr B4E0S6 B4E0S6_HUMAN	571.8	17	21,7



**Table 5. Top 47 proteins identified from Mass spec analysis of Flag IP from HEK293T-circZNF609 ORF without doxycycline induction.**

No.	Protein name	Accession	Scores	#Peptides	SC [%]
1	KIF11	sp P52732 KIF11_HUMAN	4188.7	61	50,9
2	PPM1B	sp O75688 PPM1B_HUMAN	2820.2	38	65,1
3	STK38	tr A0A024RD18 A0A024RD18_HUMAN	2588.9	44	60,4
4	STK38L	sp Q9Y2H1 ST38L_HUMAN	2455.3	36	52,4
5	RNF219	sp Q5W0B1 RN219_HUMAN	2447.5	36	40,1
6	TUBB	tr Q5SU16 Q5SU16_HUMAN	2258.0	36	65,8
7	TUBB4B	sp P68371 TBB4B_HUMAN	2189.0	38	68,3
8	TUBB2A	sp Q13885 TBB2A_HUMAN	1834.8	30	46,1
9	TUBB2A	sp P04350 TBB4A_HUMAN	1788.2	31	61,7
10	WDR77	tr A0A024R0H7 A0A024R0H7_HUMAN	1757.1	27	66,7
11	C11orf84	sp Q9BUA3 CK084_HUMAN	1626.6	21	67,7
12	RIOK1	sp Q9BRS2 RIOK1_HUMAN	1532.8	29	41,7
13	HDX	sp Q7Z353 HDX_HUMAN	1472.5	30	48,0
14	HEL-S-89n	tr V9HWB4 V9HWB4_HUMAN	1428.8	26	39,0
15	PRMT5	sp O14744 ANM5_HUMAN	1369.1	26	43,0
16	KCTD17	sp Q8N5Z5 KCD17_HUMAN	1326.3	22	52,6
17	KCTD2	sp Q14681 KCTD2_HUMAN	1317.5	12	48,7
18	N/A	tr B3KPS3 B3KPS3_HUMAN	1309.6	21	57,2
19	TUBA1C	sp Q9BQE3 TBA1C_HUMAN	1249.9	21	54,3
20	MAP3K7IP1	tr A8K6K3 A8K6K3_HUMAN	1222.2	23	59,5
21	EIF4B	sp P23588 IF4B_HUMAN	1055.4	15	25,4
22	UNC84B	tr A0A024R1P7 A0A024R1P7_HUMAN	1031.1	21	44,1
23	OGT	sp O15294 OGT1_HUMAN	1008.6	23	26,1
24	TUBB6	tr Q2NKY5 Q2NKY5_HUMAN	999.2	15	26,2
25	PABPC1	tr A0A024R9C1 A0A024R9C1_HUMAN	990.1	17	36,3

26	N/A	tr Q5U0D0 Q5U0D0_HUMAN	974.5	22	33,0
27	HSP90AB1	tr A0A024RD80 A0A024RD80_HUMAN	934.5	20	31,8
28	N/A	tr A0A140VJQ4 A0A140VJQ4_HUMAN	901.2	18	44,4
29	EL52	tr K9JA46 K9JA46_HUMAN	879.6	21	31,3
30	HNRNPH1	sp P31943 HNRH1_HUMAN	797.9	14	39,4
31	KRT5	sp P13647 K2C5_HUMAN	792.1	17	26,6
32	CLNS1A	sp P54105 ICLN_HUMAN	786.4	11	56,1
33	DHX9	sp Q08211 DHX9_HUMAN	783.1	17	17,2
34	PRMT1	tr E9PKG1 E9PKG1_HUMAN	780.3	15	45,8
35	FLJ51361	tr B4DWU6 B4DWU6_HUMAN	778.7	14	32,7
36	RBM10	tr A0A0S2Z4W4 A0A0S2Z4W4_HUMAN	751.4	18	24,6
37	DHX15	sp O43143 DHX15_HUMAN	745.4	18	23,1
38	TAB3	sp Q8N5C8 TAB3_HUMAN	729.7	14	23,5
39	HEL-S-72p	tr V9HW22 V9HW22_HUMAN	718.8	17	26,6
40	N/A	tr Q5JPK0 Q5JPK0_HUMAN	717.8	14	26,6
41	PABPC4	tr B1ANR0 B1ANR0_HUMAN	690.6	12	25,2
42	RPS3	sp P23396 RS3_HUMAN	679.7	16	65,0
43	SPIN	tr A0A024R297 A0A024R297_HUMAN	652.6	12	50,0
44	KCTD5	sp Q9NXV2 KCTD5_HUMAN	635.6	8	50,9
45	TRIM21	sp P19474 RO52_HUMAN	631.5	16	32,6
46	PRPF31	tr F1T0A5 F1T0A5_HUMAN	609.6	11	28,3
47	FLJ35376	tr B3KS36 B3KS36_HUMAN	594.9	12	25,4

## 6 Materials and methods

### 6.1 Materials

#### 6.1.1 Instruments

**Table 6: The instruments used in this study**

Screen Eraser-K	Bio-Rad (Hercules, USA)
Trans-Blot SD	Bio-Rad (Hercules, USA)
Wet-blot	Bio-Rad (Hercules, USA)
Odyssey Infrared Imaging System	LI-COR Biosciences (Lincoln, USA)
Power Supply EV233	Consort (Turnhout, Belgium)
Polymax 2040	Heidolph (Schwabach, Germany)
Vortexer REAX top	Heidolph (Schwabach, Germany)
2720 Thermal Cycler	Applied Biosystems (Foster City, USA)
GeneAmp PCR System 9700	Applied Biosystems (Foster City, USA)
Mastercycler gradient	Eppendorf (Hamburg, Germany)
Thermomixer compact	Eppendorf (Hamburg, Germany)
Centrifuge 5415D	Eppendorf (Hamburg, Germany)
Hybridization oven T 5042	Heraeus (Hanau, Germany)
Incubator Model B6200	Heraeus (Hanau, Germany)
Agilent 2100 Bioanalyzer	Agilent Technologies (Böblingen, Germany)
HeraCell 240i CO2 Incubator	Thermo Scientific (Rockford, USA)
Megafuge 40R	Thermo Scientific (Rockford, USA)
HeraSafe KS	Thermo Scientific (Rockford, USA)
Branson Sonifier 450	Heinemann (Schwäbisch Gmünd, Germany)
Film Processor CP 1000	AGFA (Mortsel, Belgium)
Geiger Counter LB123 EG&G	Berthold (Bad Wildbad, Germany)
Microscope Diavert	Leica (Wetzlar, Germany)
Milli-Q PLUS	Millipore (Billerica, USA)
Ultraspec 3300 pro	Amersham Biosciences (Little Chalfont, UK)
Avanti J-20 XP Centrifuge	Beckman Coulter (Krefeld, Germany)
Quantum ST4	PeqLab (Erlangen, Germany)
FastPrep®24 (with Lysing Matrix D)	MP Biomedicals (Santa Ana, USA)

### 6.1.2 Plasmids

**Table 7. Following plasmids were available in advance of this work or purchased**

<b>Name</b>	<b>Source</b>	<b>Purpose</b>
VP5-Flag/HA-FA (modified pIRES-neo)	AG Meister	Expression of protein N-terminally fused to Flag-HA
pCS2-myc(6)-FAME	AG Meister	Expression of protein N-terminally fused to 6xMyc
VP5-eYFP	AG Meister	Control for co-IP
PCS2-eYFP	AG Meister	Control for co-IP
pSuper Empty	AG Meister	Test expression of shRNA targeting circZNF609/circZfp609
pSuperior Puro	AG Meister	Test expression of shRNA targeting circZNF609/circZfp609
pSuperior pri-miR-331 GFP	AG Meister	Test expression of shRNA targeting circZNF609/circZfp609
pLVX Tet One Puro	AG Meister	Lentiviral plasmids for shRNA targeting circZNF609/circZfp609
TRMPV-Hygro	Addgene #27996	Cloning Venus-IRES-HygB for lentiviral vectors
pcDNA3.1 (+)- Zfp609i	AG Irene Bozzoni	Test expression of circZfp609
pcDNA3.1 (+)- Zfp609i-3x Flag	AG Irene Bozzoni	Test expression of circZfp609
pcDNA3.1(+) ZKSCAN1 MCS-WT Split GFP	Addgene #69908	Overexpression of circZfp609
pGEM-3E5-T7t	AG So Umekage	Production of synthetic circZNF609
pMD2G	AG Christoph Klein	Production of lentivirus
pspAX	AG Christoph Klein	Production of lentivirus

## 6.1.3 Oligos

**Table 8. The oligos used in this study**

No.	Name	Sequence
1	VP5-F-ciZNF609 Fw	AGTCGGCCGGCCATGTCCTTGAGCAGTGGAG
2	VP5-A-ciZNF609 Rv	GACTGGCGCGCCCTTAGAGTCAACGTCCCAC
3	q-cirZNF609-Div Fw	CAGCGCTCAATCCTTTGGGA
4	q-cirZNF609-Div Rv	GACCTGCCACATTGGTCAGTA
5	ZNF609 Fw	CTGGCTTACCCAAGGAGGTG
6	ZNF609 Rv	GTCAGGTAGTCCGGAGGTCT
7	m-ciZfp609 F1	GAAGGGGAGAATGAGTGCCG
8	m-ciZfp609 R1	GTC AAC GTC CCA CCT CAA GGT TC
9	m-mZfp609 R1	TGA CCA CTG GCA CTA ACA GG
10	sh-ciZ609 S 01	GATCTCCAGTCAAGTCTGAAAAGCAATTCAAGAGA TTGCTTTTCAGACTTGACTTTTTTA
11	sh-ciZ609 AS 01	AGCTTAAAAAAGTCAAGTCTGAAAAGCAATCTCTT GAATTGCTTTTCAGACTTGACTGGA
12	sh-ciZ609 S 02	GATCTCCTCTGAAAAGCAATGATGTTTTCAAGAGA AACATCATTGCTTTTCAGATTTTTA
13	sh-ciZ609 AS 02	AGCTTAAAAATCTGAAAAGCAATGATGTTTCTCTTG AAACATCATTGCTTTTCAGAGGA
14	sh-ciZ609 S 03	GATCTCCGTCAAGTCTGAAAAGCAATTTCAAGAGA ATTGCTTTTCAGACTTGACTTTTTA
15	sh-ciZ609 AS 03	AGCTTAAAAAGTCAAGTCTGAAAAGCAATTCTCTTG AAATTGCTTTTCAGACTTGACGGA
16	h-ciZ609-sh01-S NB	AGTCAAGTCTGAAAAGCAA
17	h-ciZ609-sh01-AS NB	TTGCTTTTCAGACTTGACT
18	h-ciZ609-sh02-S NB	TCTGAAAAGCAATGATGTT
19	h-ciZ609-sh02-AS NB	AACATCATTGCTTTTCAGA
20	h-ciZ609-sh03-S NB	CCGTCAAGTCTGAAAAGCAAT

21	h-ciZ609-sh03-AS NB	ATTGCTTTTCAGACTTGACGG
21	hZNF609 NB 01	TCAGTACATGCCCAGTGGACAACATCATTGCTTTT CAGACTTGACTTTCT
21	m-ciZNF609 NB	AGAACATGTCCAGTGGACAACATCATTGCTTTTCA GACTTGACTTTCTTTAGTG
22	h-ciZ609-sh01-F-QC	CAAACCCAGTCAAGTCTGAAAAGCAATAACCAACC TAAGCTCGCGC
23	h-ciZ609-sh01-R-QC	ATCTGAAGTCAAGTCTGAAAAGCAATAACAAACCC AAACAAAAC
24	h-ciZ609-sh02-F-QC	CAAACCCCGAGTCTGAAAAGCAATGATAACCAACC TAAGCTCGCGC
25	h-ciZ609-sh02-R-QC	ATCTGTCAAGTCTGAAAAGCAATGATAACAAACCC AAACAAAAC
26	h-ciZ609-sh03-F-QC	CAAACCCAGGTCAAGTCTGAAAAGCAAAACCAACC TAAGCTCGCGC
27	h-ciZ609-sh03-R-QC	ATCTGAAAGTCAAGTCTGAAAAGCAAAACAAACCC AAACAAAAC
28	h-ciZ609-sh01 NB	AAGTCAAGTCTGAAAAGCAAT
29	h-ciZ609-sh02 NB	TCAAGTCTGAAAAGCAATGAT
30	h-ciZ609-sh03 NB	AAAGTCAAGTCTGAAAAGCAA
31	pCDNA-ZK-ciZFP Rv	ACATGTCCAGTGGACAACATCATTGCTGAAGTATA AAAAAAAAGTCATTAG
32	pCDNA-ZK-ciZFP Fw	ACTAAAGAAAGTCAAGTCTGAAAAGGTAAGAAGCA AGGTTTCATTTAGG
33	ciZFP-pCDNA-ZK Fw	TAATGACTTTTTTTTTTATACTTCAGCAATGATGTTGT CCACTGGACATGT
34	ciZFP-pCDNA-ZK Rv	CCTAAATGAAACCTTGCTTCTTACCTTTTCAGACTT GACTTTCTTTAGT
35	p-ZK-ciZFP-3xF Rv	CACCGTCATGGTCTTTGTAGTCTTGCTGAAGTATA AAAAAAAAGTCATTAG
36	ciZFP-3xF-p-ZK Fw	CTAATGACTTTTTTTTTTATACTTCAGCAAGACTACA AAGACCATGACGGTG
43	Flag-pZK Rv	TCACTTGTCATCGTCATCCTTGTAATCG

44	5UTR-Z609-30del Fw	AGGCAGGTCTGAGAACGTAGCTGAAGC
45	5UTR-Z609-60del Fw	AAATAGGAAAGCTGGGGGCAAGGAAGAGC
46	5UTR-Z609-90del Fw	CTTGAACCTTGAGGTGGGACGTTGACTC
47	5UTR-Z609-121del Fw	ATGTCCTTGAGCAGTGGAGCCTGCGGAGGG
48	Zfp609-A1 QC Fw	GACGTTGACTCTAAGCTGTCCTTGAGCAGTG
49	Zfp609-A1 QC Rv	CACTGCTCAAGGACAGCTTAGAGTCAACGTC
50	Zfp609-A2 QC Fw	CAGCAGAACTGGAAGTGTGAGGCTCCAAGG
51	Zfp609-A2 QC Rv	CCTTGGAGCCTGACAGTTCCAGTTTCTGCTG
52	Zfp609-A1-del QC Fw	GGGACGTTGACTCTAAGTGTCTCTTGAGCAGTGGA G
53	Zfp609-A1-del QC Rv	CTCCACTGCTCAAGGACACTTAGAGTCAACGTCCC
54	hRPL32 F	AGGCATTGACAACAGGGTTC
55	hRPL32 R	GTTGCACATCAGCAGCACTT
56	FseI-h-circZNF609 F	GATAGGCCGGCCTCCTTGAGCAGTGGAGCCTCCG
57	Ascl-h-circZNF609 R	CAGTGGCGCGCCTCATTGCTTTTCAGACTTGACT
58	ZNF609 420Rv Seq	GTCTGGTAGTGTGGCCACAG
59	ZNF609 310F Seq	CCAATGCTGTGGCCACACTA
60	INTS13 FseI F	TCAAGGCCGGCCATGAAGATTTTTCTGAATCTCA
61	INTS13 Ascl R	TAGAGGCGCGCCTCACTGCCGGCTGGCTTTTCCA T
62	INTS10 FseI F	TCAAGGCCGGCCATGTCACCTAGCAAACGTAGCT CT
63	INTS10 Ascl R	TAGAGGCGCGCCTCAGGTCAGAGTCTGAAGGAGC A
64	ZK-Zfp-3ntSS-Del F	AGAAGCAAGGTTTCATTTAGGGGAA
65	ZK-Zfp-3ntSS-Del R	CTTTTCAGACTTGACTTTCTTTAGTGG
66	5UTR-Z609-121-A del Fw	TGTCCTTGAGCAGTGGAGCCTGCGGAGGG
67	5UTR-Z609-121-AT del	GTCCTTGAGCAGTGGAGCCTGCGGAGGG

	Fw	
68	5UTR-Z609-121-ATG del Fw	TCCTTGAGCAGTGGAGCCTGCGGAGGG
69	pZK-Zfp-1ATG-Flag R	ATCTTTATAATCACCGTCATGGTCTTTGTAGTCCAT CTTAGAGTCAACGTCCCACCTC
70	pZK-Zfp-1ATG-Flag F	CATGACATCGATTACAAGGATGACGATGACAAGTC CTTGAGCAGTGGAGCCTGCGGAGG
71	pGEM-T7t Fw	TAATTGAGGCCTGAGTATAAGGTGAC
72	pGEM-T7t Rv	CATTATGTTCAGATAAGGTCGTTAATC
73	h-ci-Z609-T7t-eFw	ATCTGAACATAATGCTACCCAATGATGTTGTCCACT GGGCATGTACTGAC
74	h-ci-Z609-T7t-e Rv	GTCACCTTATACTCAGGCCTCAATTAACCCCTTTTC AGACTTGACTTTCTT
75	T7t-hZ609-3xFlag-R	GATCTTTATAATCACCGTCATGGTCTTTGTAGTCTT GCATTATGTTCAGATAAGGTCG
76	T7t-hZ609-3xFlag-F	ATGACATCGATTACAAGGATGACGATGACAAGTGA TGTTGTCCACTGGGCATGTACTGACC
77	AfIII-Zeo-F	GATCCTTAAGGCTAGAGCCACCATGGCCAAGTTGA CCAGTGCCG
78	BstBI-Zeo-R	TTCGAACCCCAGAGTCCCGCTCAGTCCTGCTCCTC GGCCA
79	pTRE3G-TetOn Fw	CGGTAGAATTCTGCTTTGCATACTTCTGCCTG
80	pTRE3G-TetOn Rv	TGTCAGCAGATCTACGCGTTTTACGAGGGTAGGAA GTG
81	shRNAZ609 Rv	GTA AACGCGTAGATCTCCTGCAGGGCTGACAAC GTACAGAAGGCTCC
82	shRNAZ609 Fw	AGCTGGGTACCGTCGACTTAATTAAGATATCGGGC CCCCCTCGAGGAAGACTTTG
83	pS-IRES-Zeo Fw	ATTCTTCTGACACAACAGTCTCGAAGCCCCTCTCC CTCCCCCCCCCTAAC
84	pS-IRES-Zeo Rv	CTTGGCCATGGTGGCTCTAGCCTTAAGTGTGGCAA GCTTATCATCGTGTTTTTC
85	pLVL-Ven-IRES-HyB f	GTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTT GCAAAACGCGACCATG



86	pLVL-Ven-IRES-HyB R	TCCAGAGGTTGATTGTTCCAGACGCGTCTATTCCT TTGCCCTCGGACGAG
87	BamHI-sh609-lenti F	GTCTGGATCCGGTACCGTCGACTTAATTAAGATA
88	EcoRI-sh609-lenti R	TCGAGAATTCACGCGTAGATCTCCTGCAGGGCTG
89	pZK-Zfp-1ATG-1xHA R	GAACATCGTATGGGTACATCTTAGAGTCAACGTCC CACC
90	pZK-Zfp-1ATG-1xHA F	CTGACTATGCGGGCTCCTTGAGCAGTGGAGCCTG CGG
91	pZK-Zfp-2ATG-1xHA R	GAACATCGTATGGGTACATTTCCAGTTTCTGCTGG TCC
92	pZK-Zfp-2ATG-1xHA F	CTGACTATGCGGGCTCAGGCTCCAAGGAGGTGGG GATAC
93	3xFlag-ATG-CTG Rv	ATCTTTATAATCACCGTCGTGGTCTTTGTAGTCTTG CTGAAGTAT
94	3xFlag-ATG-CTG Fw	CACGACATCGATTACAAGGACGACGACGACAAGT GATGTTGTCCACTGGACATGTTCTGAC
95	3xFlag-3ATG-CTG Rv	ATCTTTATAATCACCGTCATGGTCTTTGTAGTCTTG CTGAAG
96	Ascl-hZNF609 R	ATATGGCGCGCCTCACCTCCTGGGGGGTGGGTAG AGT
97	q-mGapdh Fw	CTTCATTGACCTCAACTACATG
98	q-mGapdh Rv	TCGCTCCTGGAAGATGGTGATG
99	q-hGAPDH Fw	TGGTATCGTGGAAGGACTCATGAC
100	q-hGAPDH Rv	ATGCCAGTGAGCTTCCCGTTCAGC
101	Immature U1 F	GATGTGCTGACCCCTGCGATTTC
102	Immature U1 R	GTCTGTTTTTGAACTCCAGAAAGTC
103	Immature U2 F	TTGCAGTACCTCCAGGAACGG
104	Immature U2 R	CAGGGAAGCAGTTAAGTCAAGCC

#### 6.1.4 Antibodies

**Table 9. The antibodies used in this study**

<b>Antibody</b>	<b>Properties/Source</b>	<b>Dilution</b>
rat-anti-Ago2	Monoclonal, clone 11A9, Rüdél et al., 2008	1:10
rabbit-anti-GAPDH	Polyclonal, clone, clone FL-335, Santa Cruz Biotechnology	1:500
mouse-anti-p54nrb (NONO)	Monoclonal, clone 3, BD Biosciences (Franklin Lakes, USA)	1:1000
mouse-anti-Flag	Monoclonal, clone M2, Sigma Aldrich (Munich, Germany)	1:1000
mouse-anti-HA	Monoclonal, clone 16B12, Covance (Princeton, USA)	1:1000
rabbit-anti-ZNF609	AV31708, Sigma Aldrich (Munich, Germany)	1:1000
guinea pig anti-Zfp609	Polyclonal, from Debbie van den Berg et al., 2017	1:100
rabbit-anti-c-Myc	Polyclonal, C3956, Sigma Aldrich (Munich, Germany)	1:1000
mouse-anti- $\alpha$ -Tubulin	Monoclonal, clone DM1A, Sigma Aldrich (Munich, Germany)	1:10000
mouse-anti- $\beta$ -Tubulin	Abcam (Cambridge, UK)	
rabbit-anti-Lamin A/C	Polyclonal, H-110, Santa Cruz Biotechnology (Dallas, USA)	1:1000
mouse-anti- $\beta$ -Actin	Monoclonal, clone AC15, Abcam (Cambridge, UK)	1:10000
goat-anti-rabbit 680	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000
goat-anti-mouse 680	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000
goat-anti-rabbit 800	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000
goat-anti-mouse 800	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000

goat-anti-rat 800	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000
donkey-anti-guinea pig 800	secondary antibody, Li-Cor Biosciences (Lincoln, USA)	1:15000
goat-anti-rabbit 488	secondary antibody, Life Technologies (Carlsbad, USA)	1:500

### 6.1.5 Bacterial strains

Following *Escherichia coli* strain was used for cloning: XL1-blue for all plasmids, except lentiviral plasmids. Stbl3 was used for all lentiviral plasmids.

### 6.1.6 Mammalian cell lines

All cell lines are listed at ATCC: HEK293T, HCT116, HT29, CMT93. Specifically, HCT116.hCG.Luc, HT29.hCG.Luc, CMT93.hCG.Luc cell lines were provided by AG Christina Hackl as described previously<sup>142,205</sup>.

### 6.1.7 Mouse strains

Following mouse strains were used or generated during this study:

Both C57BL/6 and SCID were obtained from Charles River Laboratories (Sulzfeld)

### 6.1.8 Buffers and solutions

**Table 10. List of buffers and solutions used in this study**

Phosphate buffered saline (PBS)	130 mM NaCl
	774 mM Na <sub>2</sub> HPO <sub>4</sub>
	226 mM NaH <sub>2</sub> PO <sub>4</sub>
Tris/Borate/EDTA buffer (TBE)	89 mM Tris pH 8,3
	89 mM Boric acid
	2,5 mM EDTA
Laemmli protein sample buffer	300 mM Tris pH 6,8
	50 % Glycerol

---

	10 % SDS
	0,01 % Bromophenol blue
Tris buffered saline (TBS)	10 mM Tris pH 7,5
	150 mM NaCl
Proteinase K buffer	300 mM NaCl
	200 mM Tris pH 7,5
	25 mM EDTA
	2 % SDS
2 x RNA loading dye for Urea PAGE	99,9 % Formamide
	0,05 % Xylene cyanol
	0,05 % Bromphenol blue
20 x SSC for Northern Blot	3 M NaCl
	0,3 M NaCitrate pH 7,0
50 x Denhardt's Solution for Northern Blot	1 % Albumin Fraction V
	1 % Polyvinylpyrrolidone K30
	1 % Ficoll 400
Hybridization solution for Northern Blot	5 x SSC
	20 mM NaPi, pH 7,2
	7 % SDS
	1x Denhardt's Solution
EDC crosslinking solution	61,25 ul Methylimidazole
	75 ul 1 M HCl
	188,25 mg EDC
	Ad 6 ml H <sub>2</sub> O
Blue Juice	3 mL Glycerol (100%)

---

	200 ul 2.5% Bromphenol Blue
	6.8 mL RNase-free H <sub>2</sub> O
RNA Loading Buffer (2X) for MOPS Agarose	450 ul Formamide (100%)
	100 ul 10X MOPS
	160 ul Formaldehyde (37%)
	90 ul RNase-free H <sub>2</sub> O
	200 ul Blue Juice
Colloidal Coomassie solution	8,3 % H <sub>3</sub> PO <sub>4</sub>
	8,3 % (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
	20 % Methanol
	1 g/L Coomassie Brilliant Blue G250
MOPS-buffer (10X)	200 mM MOPS
	20 mM NaAc
	10 mM EDTA
	Dissolve in RNase-free H <sub>2</sub> O, pH 7.0
RNA-Running buffer for	1/10 MOPS buffer (10X)
MOPS Agarose (1X)	2% Formaldehyde (37%)
Wash Solution 1 for Northern Blot (Small RNA)	5x SSC
	1 % (w/v) SDS
Wash Solution 2 for Northern Blot (Small RNA)	1x SSC
	1 % (w/v) SDS
Wash Solution 3 for Northern Blot (long and circular RNA)	2X SSC
	0.1% (w/v) SDS
Wash Solution 4 for Northern Blot	0.5X SSC

---

(long and circular RNA)	0.1% (w/v) SDS
Wash Solution 5 for Northern Blot (long and circular RNA)	0.1X SSC 0.1% (w/v) SDS
SDS running buffer	200 mM Glycine 25 mM Tris pH 7,5 25 mM SDS
Towbin buffer	38,6 mM Glycine 48 mM Tris 0,0037 % (w/v) SDS 20 % Methanol
IP lysis buffer	150 mM KCl 25 mM Tris pH 7,5 2 mM EDTA 1 mM NaF 0,5 % NP-40 1mM DTT 1mM AEBSF
IP wash buffer	300 mM NaCl 50 mM Tris pH 7,5 1 mM NaF 0,01 % NP-40 5 mM MgCl <sub>2</sub>
2 x HEPES for Calcium phosphate transfection	274 mM NaCl 54,6 mM Hepes

---

	1,5 mM Na <sub>2</sub> HPO <sub>4</sub>
PBS-A for IF	0,2 g KCl
	0,2 g KH <sub>2</sub> PO <sub>4</sub>
	8 g NaCl
	2,2 g Na <sub>2</sub> HPO <sub>4</sub> _ H <sub>2</sub> O
	ad. 1 l H <sub>2</sub> O, adjust to pH 7,4
Fixation solution for IF	4 % paraformaldehyde (PFA) in PBS-A
Stopping solution for IF	100 mM Glycine in PBS-A
Permeabilization solution for IF	0,1 % Triton-X-100 in PBS-A
Wash solution for IF	5 % bovine serum albumin (BSA)
	0,01 % Triton-X-100
	ad PBS-A
Hypotonic lysis buffer (HLB)	10 mM Tris (pH 7.5)
	10 mM NaCl
	3 mM MgCl <sub>2</sub>
	0.3% (vol/vol) NP-40
	10% (vol/vol) glycerol
Nuclear lysis buffer (NLB)	20 mM Tris (pH 7.5)
	150 mM KCl
	3 mM MgCl <sub>2</sub>
	0.3% (vol/vol) NP-40
	10% (vol/vol) glycerol

## 6.2 Methods

### 6.2.1 Methods for working with DNA

#### 6.2.1.1 General molecular cloning

General methods like DNA gel electrophoresis, PCR, purification, and extraction of DNA were performed according to Sambrook et al., 1989 or as described according to manufacturer's manuals. Cloning was performed using T4 DNA ligase (Thermo Scientific) or with T4 DNA polymerase (Thermo Scientific) according to SLIC methods<sup>206</sup>.

For plasmids DNA isolation from E.coli, the NucleoSpin Plasmid (Macherey-Nagel, Düren, Germany) or NucleoBond XtraMidi-Kit (Macherey-Nagel, Düren, Germany) was used. For the elution of DNA fragments from agarose gels, the NucleoSpin Gel and PCR-cleanup (Macherey-Nagel, Düren, Germany) was used.

Sequencing has been done at Macrogen (Amsterdam, Netherlands). DNA concentration and quality were determined using NanoDrop 1000 Spectrophotometer (Thermo Scientific, Rockford, USA).

#### 6.2.1.2 List of plasmids cloned during this study

**Table 11. Details about plasmids cloned during this study**

Name	Source/plasmid backbone	Cloning oligos	Cloning enzyme	Shorted name used in the figures
VP5-Flag/HA-FA-circZNF609 ORF	VP5-Flag/HA-FA	56, 57	FseI/AscI	VP5-circZNF609 ORF
VP5-Flag/HA-FA-circZfp609 ORF	VP5-Flag/HA-FA	1, 2	FseI/AscI	VP5-circZfp609 ORF
VP5-Flag/HA-FA-ZNF609	VP5-Flag/HA-FA	96, 56	FseI/AscI	VP5-ZNF609 ORF
pCS2-myc(6)-INTS10	pCS2-myc(6x)	60, 61	FseI/AscI	6xMyc INTS10



pCS2-myc(6)- INTS13	pCS2-myc(6x)	62, 63	FseI/AscI	6xMyc INTS13
pSuper sh ZNF609 01	pSuper	10, 11	BglII, HindIII	
pSuper sh ZNF609 02	pSuper	12, 13	BglII, HindIII	
pSuper sh ZNF609 02	pSuper	14, 15	BglII, HindIII	
pSuperior CMV-sh ZNF609 01	pSuperior pri- miR-331	22, 23	Quick change	
pSuperior CMV-sh ZNF609 02	pSuperior pri- miR-331	24, 25	Quick change	
pSuperior CMV-sh ZNF609 03	pSuperior pri- miR-331	26, 26	Quick change	
pSuperior CMV- pri-miR-331 GFP- Zeocin	pSuperior CMV- pri-miR-331 GFP	77, 78	AfIII, BstBI	
pSuperior TRE-pri- miR-331 GFP- Zeocin	pSuperior CMV- pri-miR-331 GFP- Zeocin	79, 80	EcoRI, BglII	
pSuperior TRE-sh ZNF609 01 GFP- Zeocin	pSuperior TRE- pri-miR-331 GFP- Zeocin	81, 82	KpnI, MluI	
pSuperior TRE-pri- sh ZNF609 01 GFP-IRES-Zeocin	pSuperior TRE- sh ZNF609 01 GFP- Zeocin	83, 84	AfIII & SLIC	
pLVX Tet One Venus-IRES-HygB	pLVX Tet One Puro	85, 86	AvrII, MluI & SLIC	
pLVX Tet One -sh ZNF609 01 Venus-IRES-HygB	pLVX Tet One Venus-IRES- HygB	87, 88	BamHI, EcoRI	
pcDNA3.1(+)	pcDNA3.1(+)	31, 32,	SLIC	ZK-circZfp609

ZKSCAN1 circZfp609	ZKSCAN1 MCS- WT Split GFP	33, 34		
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	pcDNA3.1(+) ZKSCAN1 MCS- WT Split GFP	32, 34, 35, 36	SLIC	ZK-circZfp609- 3xFlag
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 1st AUG -> CUG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	48, 49	Quick change	1st AUG -> CUG
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 2 <sup>nd</sup> AUG -> CUG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	50, 51	Quick change	2 <sup>nd</sup> AUG -> CUG
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 1st AUG -> UG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	52, 53	Quick change	1st AUG -> UG
pcDNA3.1(+) ZKSCAN1 circZfp609 delSS	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	64, 65	Quick change	ZK-circZfp609- delSS
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del SS	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	64, 65	Quick change	ZK-circZfp609- 3xFlag-delSS
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 30	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 44	Quick change	Del 30
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 60	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 45	Quick change	Del 60

pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 89	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 46	Quick change	Del 89
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 121	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 47	Quick change	Del 121
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 121-A	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 66	Quick change	Del 121-A
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 121-AU	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 67	Quick change	Del 121-AU
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag del 121-AUG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	43, 68	Quick change	Del 121-AUG
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 4 AUG -> CUG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	93, 94	Quick change	
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 3 last AUG -> CUG	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	94, 95	Quick change	
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 1 <sup>st</sup> AUG-1xHA	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	89, 90	Quick change	

pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag 2 <sup>nd</sup> AUG-1xHA	pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag	91, 92	Quick change	
pcDNA3.1(+) ZKSCAN1 circZfp609- 3xFlag first AUG	pcDNA3.1(+) ZKSCAN1 circZfp609	69, 70	Quick change	ZK-circZfp609- 3xFlag first AUG
pGEM-3E5-T7t- apdel	pGEM-3E5-T7t	71, 72	Quick change	pGEM-3E5- T7t-apdel
pGEM-3E5-T7t- circZNF609	pGEM-3E5-T7t	71, 72, 73, 74	Quick change	circZNF609
pGEM-3E5-T7t- circZNF609 3x Flag	pGEM-3E5-T7t- circZNF609	75, 76	Quick change	circZNF609 3x Flag

## 6.2.2 Methods for working with RNA

### 6.2.2.1 RNA isolation

RNA extraction from mammalian cells has been done with TRIzol reagent (Life Technologies, Carlsbad, USA) or NucleoSpin RNA Kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol.

To isolate RNA from tissues, the mouse tissue was mechanically disrupted using FastPrep-24 (MP Biomedicals) with lysing matrix D containing 1ml TRIzol reagent. The program is set for 45 s at 6.5 m/s. Afterwards, the ceramic beads in TRIzol were incubated at RT for 5 mins and then following the TRIzol manual.

RNA extraction from the blood, serum, cell lysates or *in vitro* transcription was collected in 750 ul TRIzol-LS reagent (Life Technologies, Carlsbad, USA) following TRIzol-LS manual.

RNA isolation from MOPS-Agarose gel was performed using Zymoclean™ Gel RNA Recovery Kit (Zymo Research) following the manufacturer's protocol.

After isolation, RNA was resuspended and stored in H<sub>2</sub>O at -80°C.

#### 6.2.2.2 RNA concentration and quality control

RNA concentration was determined using NanoDrop 1000 Spectrophotometer (Thermo Scientific, Rockford, USA). For Urea PAGE, RNA quality for Northern Blot was additionally checked by nucleic acid staining of gels in TBE supplemented with 5 µl/100ml EtBr. For MOPS Agarose, RNA was stained with 2,3 µl EtBr (400µg/ml) each sample before loading into Agarose gel. RNA for deep sequencing RNAseq was analyzed using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, USA).

#### 6.2.2.3 Urea PAGE

RNA polyacrylamide gel electrophoresis was performed using the UreaGel System (National Diagnostics). Different components were mixed according to the chosen percentage and the manufacturer's protocol and gels were polymerized for at least 1 h. Afterwards, the gels were pre-run at 250-400 V for about 15-30 min using 1x TBE buffer. Samples were heated at 95°C for 2-3 min and loaded into the freshly rinsed wells. Electrophoresis was performed in 1x TBE buffer at a voltage between 250-400 V until the desired requirements.

#### 6.2.2.4 MOPS Agarose

The gel was prepared with the appropriate amount of agarose (1%) in water. After melting and cooling down to 65°C, 1x MOPS buffer and 2% Formaldehyde (37%) were added. After pouring and polymerizing, the gel was pre-run for 15-30 minutes at 70V in 1x MOPS buffer. 10 µl of 2x RNA loading dye was added to 10 µl of 20 µg of RNA samples. After incubating at 65°C for 10 minutes, samples were chilled on ice. 2.3 µl of EtBr (400 µg/ µl) were added directly into the samples before loading into the gel. For size quantification of sizes, the RNA ladder RiboRuler High Range RNA Ladder (Thermo Scientific) was used. The gel was run at 70 V until the dye approximately 1 cm above the end of the gel.

#### 6.2.2.5 Semi-Dry Blotting for Urea PAGE

In order to transfer the RNA from the Urea PAGE gel onto the nitrocellulose membrane (Amersham Hybond-N), 8 Whatman Paper with the exact measurements of the gel was prepared as well as a nitrocellulose membrane. 4 pieces of Whatman Paper were soaked in Milli-Q H<sub>2</sub>O and placed onto semi-dry blot as well as the membrane, the gel and the rest of the Whatman paper, followed by the removal of any bubbles. The blotting took place for 30-60 minutes at constant 20 V.

For crosslinking of RNA to the membrane, the EDC crosslinking solution was prepared. This solution was added to a Whatman Paper with the exact size of the membrane. The membrane was placed on top of this saturated Whatman Paper with the RNA side up, wrapped and placed into a hybridization oven at 50°C for 1 hour. Afterwards, the membrane was washed 3 times with Milli-Q H<sub>2</sub>O to remove excess crosslinking solution for 20 s each wash.

#### 6.2.2.6 Capillarity transferring for MOPS Agarose gel

Once the gel was checked under UV light, it was soaked and shaken into Milli-Q water for 5 minutes followed by 50mM NaOH for 30 minutes, 50mM Tris pH 7.5 for 30 minutes and 20xSSC for 30 minutes at room temperature. Afterwards, the transfer was set up by disposing a glass plate on top two square dishes filled with 20xSSC. Two strips of Whatman paper were soaked into 2xSSC and laid on a plate in order to hang both ends into the 20xSSC of the dishes. After removing bubbles, the gel was placed facedown. The area around the gel was sealed with parafilm to prevent 20XSSC from 'short-circuiting'. The nitrocellulose membrane (Amersham Hybond-N) of the size of the gel was soaked into Milli-Q water and 2xSSC and it was placed on the gel. After removing all bubbles, three 2xSSC-soaked pieces of Whatman paper were placed on the membrane. A stack of towel papers was put on top with a second glass plate and an additional weight of 500 g. The transfer was set at room temperature overnight, preferably for 18 hours, after which the success of transfer was verified by visualizing the membrane under UV light. The RNA was crosslinked to the membrane using the UV Stratalinker at 254 nm.

#### 6.2.2.7 Prehybridization of the membrane

For prehybridization of the membrane, 20-30 ml Hybridization Solution was put into a cylindrical glass bottle, in which the membrane was already put into with the RNA facing inwards, and then placed in a hybridization oven at 42-60°C for 60 minutes.

#### 6.2.2.8 Preparation of probes

The DNA probes were used for the detection of small RNAs, overexpressed RNAs or abundant long RNAs. There are single-stranded DNA oligonucleotides, about 50 nts. Probe labeling was done using 20 pmol DNA oligonucleotide with 20 µCi of γ 32P-ATP (Hartmann Analytics, Braunschweig, Germany) using T4 PNK enzyme (Life Technologies, Carlsbad, USA) according to the manufacturer's protocol. The reaction was incubated for 60 min at 37°C. The reaction was mixed with 30 µl 30 mM EDTA to stop PNK reaction.

The cDNA probes were prepared using Megaprime DNA-Labeling Systems kit (Amersham). Afterward, the reaction was stopped with 5  $\mu$ l 200 mM EDTA.

The DNA or cDNA probes were purified using Illustra MicroSpin G-25 columns (GE Healthcare, Little Chalfont, UK) and flow through are collected. The DNA probes are directly added to the prehybridized membranes. The cDNA probes are denatured by heating them to 95°C for 5 minutes and then immediately placed on ice. Then, the cDNA probes were added to the prehybridized membranes.

#### 6.2.2.9 Northern blot wash and detection

The membrane labeled with DNA probes was washed twice with Northern Blot wash solution 1 and once with Northern Blot wash solution 2, incubating each wash step for 10 mins on a turning wheel at the same temperature as prehybridization and hybridization.

The membrane labeled with cDNA probes was washed once with Northern Blot wash solution 3, once with Northern Blot wash solution 4 and once with Northern Blot wash solution 5, incubating each wash step for 30 mins on a turning wheel at the same temperature as prehybridization and hybridization.

After the last step, the liquid was discarded and the membrane was wrapped in saran for exposure to a phosphor screen and then signals were scanned by Personal Molecular Imager System (Bio-Rad, Hercules, USA).

#### 6.2.2.10 Stripping of Northern blot membranes

In order to reuse the membrane, the membranes were stripped off the current probe. Water was boiled, then an SDS solution was carefully added to reach a final concentration of 0.1% SDS. The membrane was rolled up with the RNA facing inwards. Repeat this step 2 times. After a final wash with only boiled water, the membrane was wrapped in saran and exposed to a screen for at least over night to control the removal of the probe.

#### 6.2.2.11 First strand cDNA synthesis

For qPCR, RNA was preferably isolated using NucleoSpin RNA Kit (Macherey-Nagel, Düren, Germany) including DNA digestion step. 1  $\mu$ g RNA was used for cDNA synthesis using First Strand cDNA Synthesis Kit (Thermo Scientific, Rockford, USA) according to its manual.

#### 6.2.2.12 Quantitative real-time PCR (qPCR)

qPCR was done with Sso Fast Eva Green Mix (Bio-Rad, Hercules, USA) using 0,5  $\mu$ M forward and 0,5  $\mu$ M reverse primer and cDNA from 10 ng RNA as template. qPCR was run on a CFX96 cyclor (Bio-Rad, Hercules, USA) using the standard program as given in the SsoFast EvaGreen SuperMix manual. Data were evaluated using the  $\Delta\Delta C_t$  method with GAPDH or RPL32 as reference mRNA and normalized to control sample.

#### 6.2.2.13 RNA *in vitro* transcription using T7 polymerase

The reaction requires a DNA template containing the T7 promoter upstream of the sequence and the corresponding T7 RNA polymerase (homemade from AG Meister). In order to perform the *in vitro* transcription, 2  $\mu$ g of DNA templates was prepared with 1x T7 buffer, 5 mM NTPs, 5mM MgCl<sub>2</sub>, TIPP, Ribolock (Thermo Scientific) and 25  $\mu$ l of T7 polymerase (2  $\mu$ g/ $\mu$ l) in DEPC-treated water in a final volume of 500  $\mu$ l. The *in vitro* transcription was performed at 37°C for 3 hours. For increasing of self-splicing reaction of circRNA, the reaction was shifted to 55°C for 30 mins. Afterwards, the reaction was digested with 1  $\mu$ l of 1U/unit DNase I (Thermo Scientific) for 15 mins at 37°C.

The RNA reaction was then purified using TRIzol-LS reagents.

#### 6.2.2.14 RNA treatment with RNase R

RNA was treated with RNase R (Biozym) according to manufacturer's protocol. The reaction was cleaned up using NucleoSpin RNA Kit (Macherey-Nagel, Düren, Germany).

#### 6.2.2.15 Poly-(A) (+/-) isolation from total RNA

mRNA was isolated using Dynabeads™ Oligo(dT)<sub>25</sub> (Thermo Scientific) according to manufacturer's protocol. The flow-through was collected and served as poly A (–) RNA. The RNA after isolation was collected and cleaned up from the magnetic beads using NucleoSpin RNA Kit (Macherey-Nagel, Düren, Germany).

### 6.2.3 Methods for working with proteins

#### 6.2.3.1 Preparation of whole cell extracts from mammalian cells

Cell extracts were prepared using IP lysis buffer. About 4 million cells were thoroughly resuspended in 100  $\mu$ l IP lysis buffer by pipetting and incubated on ice for 20 min. Lysates were then cleared by centrifugation at full speed for 15 min at 4°C. The supernatant was



supplemented with Laemmli buffer or LDS buffer (for NuPAGE® 4- 12% Bis-Tris Gel) or used for IP.

#### 6.2.3.2 Preparation of whole cell extracts from tissues

The mouse tissue was mechanically disrupted using FastPrep®-24 (MP Biomedicals) with lysing matrix D containing 200-600 µl IP lysis buffer. The program is set for 45 s at 6.5 m/s. Afterward, the lysate was collected and was cleared at full speed for 15 min at 4°C. The supernatant was collected and supplemented with Laemmli buffer.

#### 6.2.3.3 Preparation of subcellular extracts.

Nucleo-cytoplasmic fractionations were prepared according to Nature Protocols as described in Gagnon et al. 2014 with slight modifications. HEK293T cells were cultured up to 80 % confluency, washed with PBS and harvested by trypsin. The cells were then passed through a cell strainer and washed again with ice-cold PBS. After centrifugation at 100 g for 5 min at 4°C, the cell pellet was resuspended by gentle pipetting with ice-cold hypotonic lysis buffer (HLB) and incubated for 8 min on ice. 1ml of HLB was used for every 75 mg cell pellet (or 10 million cells). Afterwards, cells were centrifuged at 800g for 8 mins at 4°C. The supernatant of this step served as cytoplasmic fraction. The nuclei pellet was washed gently for 4 times with HLB through pipetting and centrifugation at 200 g for 2 mins at 4°C. Nuclei were resuspended nuclear lysis buffer (NLB). 0,5 ml of NLB was used for every 75mg of initial cells used (or 10 million cells). If needed, the nuclear pellet was sonicated on ice one time with 10-20 % power for 15s. Both fractions were cleared by centrifuge at 15.000 g for 15 mins at 4°C.

#### 6.2.3.4 Determination of protein concentrations

For the determination of protein concentrations in solutions containing a mixture of several different proteins (e.g. cellular lysates), a Bradford Protein Assay (Biorad) was used and performed according to the protocol of the distributor and as described in (Bradford, 1976)

#### 6.2.3.5 Immunoprecipitation (IP)

Anti-FLAG M2 agarose beads (Sigma-Aldrich) was used for IP of FLAG/HA-tagged proteins that were overexpressed in HEK293T cells. Prior to use, the matrix was washed with cold PBS twice. First of all, after preparing the total cell lysates, input samples were collected, supplemented with LDS sample buffer, heated at 70°C for 15 mins and stored at -20°C. For immunoprecipitation, the lysate was incubated with the antibody-coupled

beads for at 2-3 h while rotating at 4°C. After incubation, the beads were sedimented by centrifugation for 1 min at 1,000 g and the supernatant was removed. The beads were washed with the IP wash buffer four times. During the last washing step, the samples were transferred to new reaction tubes to minimize contamination by unspecific protein binding to the tube material. After a final washing step with PBS, the supernatant was removed quantitatively and the beads were eluted with Flag peptides.

#### 6.2.3.6 SDS-PAGE and Western blotting

NuPAGE® 4-12% Bis-Tris Gel (Invitrogen, Life Technologies) was used for samples prepared for Mass Spec. The proteins were visualized with Coomassie.

For performing a Western blot, samples were shortly heated at 95°C for 5 mins and then loaded onto a 10-15% SDS-PAGE gel. After separation, the proteins were blotted onto an Amersham Hybond-ECL membrane (GE Healthcare) using Towbin buffer for 1 min/kDa constant at 1 mA/cm<sup>2</sup>. The membrane was blocked in TBS containing 0.1% Tween20 and 5% milk powder and subsequently incubated with the primary antibody overnight at 4°C. After three washing steps with TBS containing 0.1% Tween 20, the secondary antibody was added for 1 hour at RT. After three washing steps with TBS containing 0.1% Tween 20, the membrane was scanned on a LI-COR reader.

#### 6.2.3.7 Mass-spectrometric analyses

After gradient-gels or SDS gels were destained, bands or gel parts were excised and transferred into 2ml micro tubes (Eppendorf), washed for 30 min with 500 µl 50 mM NH<sub>4</sub>HCO<sub>3</sub>, 50 mM NH<sub>4</sub>HCO<sub>3</sub>/ acetonitrile (3/1), 10 mM NH<sub>4</sub>HCO<sub>3</sub>/ acetonitrile (3/1), 10 mM NH<sub>4</sub>HCO<sub>3</sub>/ acetonitrile (1/1) and lyophilized. After reduction and alkylation of cysteines with 100 µl 1mg/ml DTT (57°C/ 35 min) and 200 µl 5 mg/ml Iodoacetamide (RT/ 35 min) solved in 50 mM NH<sub>4</sub>HCO<sub>3</sub>, gel slices were washed again and lyophilized. Proteins were subjected to in-gel tryptic digest overnight at 37°C with 0,8 µg Trypsin Gold mass spectrometry grade (Promega) per sample. Peptides were first extracted twice with 100 mM NH<sub>4</sub>HCO<sub>3</sub>, followed by 100 mM NH<sub>4</sub>HCO<sub>3</sub>/ acetonitrile (2/1) and eluates were combined and lyophilized. Further processing was executed by Dr. Astrid Bruckmann or Eduard Hochmuth.

## 6.2.4 Methods for working with cells

### 6.2.4.1 Working with bacterial cells

For transformations, chemically competent *E. coli* XL1-blue was used for all plasmids except lentiviral plasmids. For lentiviral vectors, Stbl3 was used. Briefly, appropriate amounts of plasmids or ligation or SLIC products were used for transformations. After adding the plasmids to the bacterial cells, the cells were incubated on ice for at least 15 min and subjected to a 60 s heat shock at 42°C. After 2 mins incubation on ice, 1 ml of LB medium without antibiotic was added, and the cells were incubated at 37°C for at least 20 min while shaking. Appropriate amounts of LB containing bacteria were plated on LB plates containing suitable antibiotic. LB plates for XL1-Blue were incubated overnight at 37°C. LB plates for Stbl3 were incubated overnight at 30-32°C.

The colonies are picked and incubated overnight in LB medium containing suitable antibiotics at the desired temperature.

### 6.2.4.2 Working with mammalian cells

All cell lines were cultivated at 37°C in a humidified chamber with air atmosphere and addition of 5 % CO<sub>2</sub> on uncoated plastic surface. All cell lines were cultured with DMEM media supplemented with 10% FCS, 1% PS.

#### 6.2.4.2.1 Cell transfection with calcium phosphate

Only HEK 293T cells were transfected using calcium phosphate. Per 15 cm plate, 10 µg DNA and 123 µl of 2 M CaCl<sub>2</sub> were filled up to 1 ml with water in a reaction tube. Afterwards, 1 ml of 2x HEPES were added dropwise into the DNA-containing solution. The mixture was incubated for 10 minutes at room temperature and then added to the cells. The cells were harvested after 24-28 hours. Per 10 cm plate, half of the reagents were used.

#### 6.2.4.2.2 Cell transfection with Lipofectamine 2000 Reagent

2ml of 400.000 cells/ml were seeded into 6-wells plates. 500 ng of *in vitro* transcribed RNA was added to 250 µl of Opti-MEM. In a second tube, 10 µl of Lipofectamine 2000 was added into 250 µl of Opti-MEM. The mixture was incubated at RT for 5 mins. After incubation, both tubes were mixed together and incubated at room temperature for further 15 minutes. 500 µl of the mixture was added drop-by-drop to the cells. 4 hours post-transfection, the medium was removed and replaced with 2ml of fresh DMEM with 10%

FBS and 1% PS. The cells were incubated for 16 hours. RNA was harvested 4 hours and 16 hours post-transfection.

For the transfection of plasmids, the same procedure was applied as with *in vitro* transcribed RNA and in this case, 2,5 µg of plasmids were used. For transfection in other culture disk format, the reagents are scaled according to manufacturer's recommendations.

#### 6.2.4.2.3 Cell transfection with Lipofectamine RNAiMAX

For siRNA, the transfection was performed as described for Lipofectamine 2000 Reagent with some modification. For example, the same amount of cells and Opti-MEM were used in 6-wells plates. In this scale, 10 µl of Lipofectamine RNAiMAX was used and a desired amount of siRNA was used (instead of 500 ng *in vitro* transcribed RNA as above). The cells were harvested after 72-96 hours.

#### 6.2.4.2.4 Generation of stable cell lines via Lipofectamine 2000 transfection

For generation of stable HEK293T, HCT116.hCG.Luc or CMT93.hCG.Luc cell line expressing desired constructs, the cells were transfected with plasmids in 6-well format. Cells were split one day post-transfection into one 10 cm plate and selection was started with 100 µg/ml Zeocin. The cells are sorted for GFP-expressing cells using FACS Aria system (BD Biosciences, Franklin Lakes, USA). Stable clones were maintained in the same medium as used for selection.

#### 6.2.4.2.5 Generation of stable cell lines via lentiviral transduction

Lentivirus was generated at the S2 lab (AG Christoph Klein). Briefly, the HEK293T cells were propagated in DMEM media supplemented with 10% FCS, 1% PS and L-glutamine.  $6 \times 10^6$  cells were seeded in a 10 cm cell culture dishes 16h before transfection. Transfection medium (DMEM; 10% FCS) containing 25 µM chloroquine was added 45 min before the start of transfection. Afterward, chloroquine was replaced by transfection mixture.

5 µg pMD2G (envelope plasmid), 20 µg psPAX2 (packaging plasmid), 20 µg lentiviral plasmid were used for preparation of the mix for calcium phosphate transfection. The transfection mixture was added dropwise to HEK293T cells. The medium was replaced 6 hours post-transfection. The viral particles were collected 48h and 72h post-transfection, filtered, concentrated and stored at -80°C for at least 6 months.

The HT29.hCG.Luc, HCT116.hCG.Luc, CMT93.hCG.Luc cells ( $4 \times 10^4$  cells/well) were plated and propagated in 6 well-plates overnight at 37°C. On the next day, after changing to the new medium virus dilution of 1:10, 1:100, 1:1000, 1:10000 was added to the cells. After 24 hours, 100 µg/ml Hygromycin B was added for lentiviral transduction cell lines. The cells were selected in Hygromycin B for 3 passages. Afterward, all the cell lines can be moved to the S1 lab. The cells are sorted for GFP-expressing cells using FACS Aria system (BD Biosciences, Franklin Lakes, USA). Stable clones were maintained in the same medium as used for selection.

#### 6.2.4.2.6 Proliferation assay

Single cell suspensions were obtained after trypsinization and passing through a cell strainer. XTT were performed in 96 well plates using XTT colorimetric assay kit (Roche) based on manufacturers instruction. Specifically, 300 cells/well for HT29.hCG.Luc or HCT116.hCG.Luc cell lines or 100 cells/well for CMT93.hCG.Luc cell lines were seeded in 12 replications. 6 hours after seeding, 6 replications were replaced with DMEM media supplemented with 10% FCS, 1% PS while the other 6 replications were replaced with DMEM media supplemented with 10% FCS, 1% PS, 1 µg/ml Doxycycline. The XTT are assayed every day for 6-8 days. Statistics analysis was performed using GraphPad Prism.

#### 6.2.4.2.7 Preparation for cell sorting

Cells were dissociated with trypsin, washed once in DMEM supplemented with 10% FCS, 1% PS. Single cells were obtained by passing through a cell strainer and then resuspended to the desired concentration, preferably 20 million cells/ml, in PBS.

#### 6.2.4.2.8 Immunofluorescence

The cells were seeded in 24-well plates with 12 mm diameter cover slips of thickness #1,5 (0,15-0,19 mm) for confocal microscopy (Carl Roth, Karlsruhe, Germany). The stable cell lines HT29.hCG.Luc or HCT116.hCG.Luc expressing shRNA targeting circZNF609 were treated with 1µg/ml Doxycycline for at least 4 days before staining. For overexpression, the cells are fixed 2 days post-transfection.

PBS, fixation solution and stopping solution were prewarmed to 37°C before addition to cells to prevent damage to cell morphology prior to fixation. All steps were done in 24-well format with 500 µl volumes for all washing steps or 300 µl for antibody-containing

solutions. Cells were washed once with PBS, fixed with fixation solution for 10 mins at 37°C and stopped with stopping solution for 5 mins.

Then, the cells were washed once with PBS and permeabilized with permeabilization solution for 30 mins at room temperature. After washing off detergent once with PBS, cells were blocked with blocking solution containing 5 % BSA for 1 hour at room temperature. Primary antibody was incubated in blocking solution for 1 h at room temperature or overnight at 4°C. After three washing steps with blocking solution and incubation for 5 min at RT between each step, the secondary antibody was added and incubated for 1 h at room temperature. Cells were then washed once with blocking solution and three times with PBS and incubated for 2 min between each wash. Afterward, PBS is removed by water. Coverslips were then mounted using 8 µl Prolong Gold with DAPI (Life Technologies, Carlsbad, USA) and dried for minimum 12 h.

Confocal microscopy has been done with a TCS SP8 (Leica Microsystems, Mannheim, Germany) inverted microscope equipped with an acousto-optical beam splitter, 405 nm laser (for DAPI), Argon laser (488 nm for Alexa 488 and GFP) and DPSS laser 561 nm (for Alexa 555).

## 6.2.5 Methods for working with animals

### 6.2.5.1 Cell preparation for injection

Cells were dissociated with trypsin, washed once in DMEM supplemented with 10% FCS, 1% PS. Single cells were obtained by passing through a cell strainer and then resuspended to the desired concentration in Hank's Balanced Salt Solution (HBSS). All animals receiving tumor cell injections were injected with the same quantity of tumor cells (0,5 million cells per animal)

### 6.2.5.2 Tumor implantation

All experimental procedures were conducted in accordance with German laws governing animal care. All tumor implantation was performed by Dr. Christina Hackl with the assistance from a technician Marvin Anders.

#### 6.2.5.3 Subcutaneous implantation

50 µl of 0,5 million tumor cells were injected subcutaneously into C57BL/6 mice using a manual 1 ml Hamilton syringe and 30G needle (BD Biosciences). Tumor size was measured daily (normally 5 days after implantation) with calipers and tumor volume calculated as width x width x length as described<sup>142</sup>.

#### 6.2.5.4 Intrasplenic implantation

Intrasplenic implantation was performed with SCID mice. Abdominal access was obtained via a left subcostal skin and peritoneal wall incision. The spleen was gently exteriorized. 5 µl of 0,5 million tumor cells were injected with a 10 µl Hamilton syringe (BD Biosciences). The needle was slowly retracted and the injection site pressed with a moist cotton swab to prevent leakage. The spleen was returned to the peritoneal cavity; peritoneum and skin were closed by running sutures and wound clips.

#### 6.2.5.5 Bioluminescence measurement

Bioluminescence measurement was performed weekly or every 3 days. For this purpose, mice were anaesthetized, injected with 100 µl of 30 mg/ml D-luciferin (Biosynth, Staad, Switzerland) for 3 minutes, and emitted photons were registered for 3 min using Xenogen IVIS Lumina Imaging System (Caliper Life Sciences, Hopkinton, USA).

#### 6.2.5.6 Doxycycline treatment

Doxycycline was obtained from Sigma Aldrich (Munich, Germany). The mice were gavaged daily with 100 µl of 5 mg/ml Doxycycline as described<sup>207</sup>.

#### 6.2.5.7 Tumor resections

Primary tumor (either subcutaneous tumor or at spleen) and metastases in the liver resections were performed when then tumor size is big enough or when the bioluminescence measurement is significant. Samples of primary tumors and liver metastasis were weighted, imaged with a normal camera. All the tumors are divided into different parts for RNA, protein or immunohistochemistry analysis. Statistics analysis was performed using GraphPad Prism

## 6.2.6 Next generation sequencing and data analysis

### 6.2.6.1 Generation of small RNA libraries

Isolated RNA was ligated to an adenylated 3' adapter by a truncated T4 RNA Ligase 2 (expressed and purified from AG Meister), the 5' RNA adapter was added in a second ligation step by T4 RNA Ligase 1 (NEB). The product was reverse-transcribed using the SuperScriptIII First Strand Synthesis Super Mix (Invitrogen) using a specific primer, followed by a PCR amplification. The samples were run on a 6% Urea-PAGE, the bands corresponding to small RNA containing ligation products were cut out and eluted overnight in elution buffer (300 mM NaCl, 2 mM EDTA). The libraries were precipitated with ethanol overnight at -20°C, then collected by centrifugation and solved in water. The library was sequenced on a MiSeq (Illumina) in a 1x66 bp run.

### 6.2.6.2 Analysis of small RNA data

Data analysis was performed using in-house written scripts. Sequences were mapped, without any mismatches allowed, against human miRNAs listed in the miRBase v20 (June 2013; <http://www.mirbase.org>). The minimum length of reads was set to 18 nucleotides. Annotated miRNA reads were normalized as RPM values according to the total number of mapped reads in the respective library.

### 6.2.6.3 Generation of rRNA-depleted RNA libraries

The rRNA-depleted RNA library was prepared using Ovation Human FFPE RNA-Seq Library System according to manufacturer's protocol. Specifically, 100 ng of total RNA from frozen samples was used.

### 6.2.6.4 Bioinformatics analysis of circRNAs

Bioinformatics analysis of circRNAs was performed by Petar Glazar from the Rajewsky lab. Specifically, circRNAs were detected and annotated as described before<sup>26</sup>. Genome references used for all mapping and subsequent analyses were human hg19 (Feb 2009, GRCh37). Quantification of circRNAs expression from sequencing data was performed as described<sup>65</sup>.



## 7 References

1. Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J. & Kleinschmidt, A. K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci.* **73**, 3852–3856 (1976).
2. Gross, H. J. *et al.* Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* **273**, 203–208 (1978).
3. Arnberg, A. C., Van Ommen, G. J. B., Grivell, L. A., Van Bruggen, E. F. J. & Borst, P. Some yeast mitochondrial RNAs are circular. *Cell* **19**, 313–319 (1980).
4. Hensgens, L. A. M. *et al.* Variation, transcription and circular RNAs of the mitochondrial gene for subunit I of cytochrome c oxidase. *J. Mol. Biol.* **164**, 35–58 (1983).
5. Grabowski, P. J., Zaug, A. J. & Cech, T. R. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of tetrahymena. *Cell* **23**, 467–476 (1981).
6. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
7. Zaug, A. J., Grabowski, P. J. & Cech, T. R. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage–ligation reaction. *Nature* **301**, 578–583 (1983).
8. Kos, A., Dijkema, R., Arnberg, a C., van der Meide, P. H. & Schellekens, H. The hepatitis delta (delta) virus possesses a circular RNA. *Nature* **323**, 558–60 (1986).
9. Nigro, J. M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
10. Cocquerelle, C., Daubersies, P., Majérus, M. a, Kerckaert, J. P. & Bailleul, B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* **11**, 1095–1098 (1992).
11. Cocquerelle, C., Mascrez, B., Héтуin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* **7**, 155–160 (1993).
12. Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
13. Li, X. F. & Lytton, J. A circularized sodium-calcium exchanger exon 2 transcript. *J. Biol. Chem.* **274**, 8153–8160 (1999).
14. Zaphiropoulos, P. G. Differential expression of cytochrome P450 2C24 transcripts in rat kidney and prostate: evidence indicative of alternative and possibly trans splicing events. *Biochem Biophys Res Commun* **192**, 778–786 (1993).
15. Zaphiropoulos, P. G. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6536–

- 6541 (1996).
16. Zaphiropoulos, P. G. Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol. Cell. Biol.* **17**, 2985–2993 (1997).
  17. Surono, a *et al.* Circular dystrophin RNAs consisting of exons that were skipped by alternative splicing. *Hum. Mol. Genet.* **8**, 493–500 (1999).
  18. Bailleul, B. During in vivo maturation of eukaryotic nuclear mRNA, splicing yields excised exon circles. *Nucleic Acids Res.* **24**, 1015–1019 (1996).
  19. Schindewolf, C. A. & Domdey, H. Splicing of a circular yeast pre-mRNA in vitro. *Nucleic Acids Res.* **23**, 1133–1139 (1995).
  20. Schindewolf, C., Braun, S. & Domdey, H. In vitro generation of a circular exon from a linear pre-mRNA transcript. *Nucleic Acids Res.* **24**, 1260–1266 (1996).
  21. Braun, S., Domdey, H. & Wiebauer, K. Inverse splicing of a discontinuous pre-mRNA intron generates a circular exon in a HeLa cell nuclear extract. *Nucleic Acids Res.* **24**, 4152–4157 (1996).
  22. Pasman, Z., Been, M. D. & Garcia-Blanco, M. A. Exon circularization in mammalian nuclear extracts. *RNA* **2**, 603–10 (1996).
  23. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, (2012).
  24. Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**, 3131–3142 (2012).
  25. Salzman, J. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* **9**, e90859 (2014).
  26. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
  27. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genet.* **9**, (2013).
  28. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
  29. Maass, P. G. *et al.* A map of human circular RNAs in clinically relevant tissues. *Journal of Molecular Medicine* 1–11 (2017). doi:10.1007/s00109-017-1582-9
  30. Szabo, L. & Salzman, J. Detecting circular RNAs: bioinformatic and experimental challenges. *Nat. Rev. Genet.* **17**, 679–692 (2016).
  31. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **32**, 453–61 (2014).

32. Zeng, X., Lin, W., Guo, M. & Zou, Q. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Computational Biology* **13**, (2017).
33. Ebbesen, K. K., Kjems, J. & Hansen, T. B. Circular RNAs: Identification, biogenesis and function. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1859**, 163–168 (2016).
34. Hansen, T. B., Venø, M. T., Damgaard, C. K. & Kjems, J. Comparison of circular RNA prediction tools. *Nucleic Acids Res.* **44**, (2015).
35. Suzuki, H. *et al.* Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34**, (2006).
36. Suzuki, H. & Tsukahara, T. A view of pre-mRNA splicing from RNase R resistant RNAs. *International Journal of Molecular Sciences* **15**, 9331–9342 (2014).
37. Gardner, E. J., Nizami, Z. F., Conover Talbot, J. & Gall, J. G. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev.* **26**, 2550–2559 (2012).
38. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **32**, 453–461 (2014).
39. Andreeva, K. & Cooper, N. G. F. Circular RNAs: New Players in Gene Regulation. *Adv. Biosci. Biotechnol.* **6**, 433–441 (2015).
40. Chen, I., Chen, C. Y. & Chuang, T. J. Biogenesis, identification, and function of exonic circular RNAs. *Wiley Interdisciplinary Reviews: RNA* **6**, 563–579 (2015).
41. Lasda, E. & Parker, R. O. Y. Circular RNAs: diversity of form and function. *Rna* 1829–1842 (2014). doi:10.1261/rna.047126.114.DIFFERENT
42. Zhang, Y. *et al.* Circular Intronic Long Noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
43. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* (2015). doi:10.1038/nsmb.2959
44. Ivanov, A. *et al.* Analysis of Intron Sequences Reveals Hallmarks of Circular RNA Biogenesis in Animals. *Cell Rep.* **10**, 170–177 (2015).
45. Ashwal-Fluss, R. *et al.* CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).
46. Chen, L. L. & Yang, L. Regulation of circRNA biogenesis. *RNA Biol.* **12**, 381–388 (2015).
47. Li, X. *et al.* Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection. *Mol. Cell* **67**, 214–227.e7 (2017).
48. Zhang, Y. *et al.* The Biogenesis of Nascent Circular RNAs. *Cell Rep.* **15**, 611–624 (2016).

49. Zhang, X. O. *et al.* Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147 (2014).
50. Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
51. Liang, D. & Wilusz, J. E. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* **28**, 2233–2247 (2014).
52. Kramer, M. C. *et al.* Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* **29**, 2168–2182 (2015).
53. Starke, S. *et al.* Exon circularization requires canonical splice signals. *Cell Rep.* **10**, 103–111 (2015).
54. Chen, L.-L. The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* **17**, 205–211 (2016).
55. Wang, Y. & Wang, Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172–179 (2015).
56. Liang, D. *et al.* The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Molecular Cell* (2017). doi:10.1016/j.molcel.2017.10.034
57. Aktaş, T. *et al.* DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* **544**, 115–119 (2017).
58. Errichelli, L. *et al.* FUS affects circular RNA expression in murine embryonic stem cell-derived motor neurons. *Nat. Commun.* **8**, 14741 (2017).
59. Fei, T. *et al.* Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc. Natl. Acad. Sci.* 201617467 (2017). doi:10.1073/pnas.1617467114
60. Khan, M. A. F. *et al.* RBM20 Regulates Circular RNA Production from the Titin Gene. *Circ. Res.* **119**, 996–1003 (2016).
61. Dubin, R. A., Kazmi, M. A. & Ostrer, H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245–248 (1995).
62. Zheng, Q. *et al.* Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.* **7**, 11215 (2016).
63. Zhang, X. O. *et al.* Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).
64. Gao, Y. *et al.* Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* **7**, 12060 (2016).
65. Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**, 870–885 (2014).

- 
66. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352 (2014).
  67. Thomson, D. W. & Dinger, M. E. Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* **17**, 272–283 (2016).
  68. Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99–110 (2011).
  69. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**, 409 (2014).
  70. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–8 (2013).
  71. Piwecka, M. *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* (80-. ). eaam8526 (2017). doi:10.1126/science.aam8526
  72. Mullokandov, G. *et al.* High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* **9**, 840–846 (2012).
  73. Bosson, A. D., Zamudio, J. R. & Sharp, P. A. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol. Cell* **56**, 347–359 (2014).
  74. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
  75. Kjems, J. & Garrett, R. A. An intron in the 23S ribosomal gene of the archaebacterium *Desulfurococcus mobilis*. *Nature* **318**, 675–677 (1985).
  76. Kjems, J. & Garrett, R. A. Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium *desulfurococcus mobilis*. *Cell* **54**, 693–703 (1988).
  77. Dalgaard, J. Z., Garrett, R. A. & Belfort, M. A Site-Specific Endonuclease Encoded by a Typical Archaeal Intron. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5414–5417 (1993).
  78. Dalgaard, J. Z. & Garrett, R. A. Protein-coding introns from the 23S rRNA-encoding gene form stable circles in the hyperthermophilic archaeon *Pyrobaculum organotrophum*. *Gene* **121**, 103–110 (1992).
  79. Chen, C. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* (80-. ). **268**, 415–417 (1995).
  80. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* (80-. ). **324**, 218–223 (2009).
  81. Oh, E. *et al.* Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308 (2011).

82. You, X. *et al.* Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat. Neurosci.* (2015). doi:10.1038/nn.3975
83. Schneider, T. *et al.* CircRNA-protein complexes: IMP3 protein component defines subfamily of circRNPs. *Sci. Rep.* **6**, 31313 (2016).
84. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res.* **27**, 626–641 (2017).
85. AbouHaidar, M. G., Venkataraman, S., Golshani, A., Liu, B. & Ahmad, T. Novel coding, translation, and gene expression of a replicating covalently closed circular RNA of 220 nt. *Proc. Natl. Acad. Sci.* **111**, 14542–14547 (2014).
86. Legnini, I. *et al.* Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **66**, 22–37.e9 (2017).
87. Pamudurti, N. R. *et al.* Translation of CircRNAs. *Mol. Cell* **66**, 9–21.e7 (2017).
88. Yang, Y. *et al.* Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* **110**, (2018).
89. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745 (2009).
90. Komar, A. A. & Hatzoglou, M. Cellular IRES-mediated translation: The war of ITAFs in pathophysiological states. *Cell Cycle* **10**, 229–240 (2011).
91. Stoneley, M. & Willis, A. E. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene* **23**, 3200–3207 (2004).
92. Meyer, K. D. *et al.* 5' UTR m6A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010 (2015).
93. Zhou, J. *et al.* Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
94. Perriman, R. & Ares, M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA* **4**, 1047–54 (1998).
95. Abe, N. *et al.* Rolling circle amplification in a prokaryotic translation system using small circular RNA. *Angew. Chemie - Int. Ed.* **52**, 7004–7008 (2013).
96. Abe, N. *et al.* Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **5**, 16435 (2015).
97. Granados-Riveron, J. T. & Aquino-Jarquin, G. The complexity of the translation ability of circRNAs. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1859**, 1245–1251 (2016).
98. Zhou, C. *et al.* Genome-Wide Maps of m6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs. *Cell Rep.* **20**, 2262–2276 (2017).

- 
99. Dudekula, D. B. *et al.* Circinteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.* **13**, 34–42 (2016).
  100. Li, B. *et al.* Discovering the Interactions between Circular RNAs and RNA-binding Proteins from CLIP-seq Data using circScan. *bioRxiv* 115980 (2017). doi:10.1101/115980
  101. Chen, X. *et al.* circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci. Rep.* **6**, 34985 (2016).
  102. Meng, X., Chen, Q., Zhang, P. & Chen, M. CircPro: An integrated tool for the identification of circRNAs with protein-coding potential. *Bioinformatics* **33**, 3314–3316 (2017).
  103. Conn, V. M. *et al.* A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* **3**, 17053 (2017).
  104. Du, W. W. *et al.* Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur. Heart J.* **38**, 1402–1412 (2017).
  105. Du, W. W. *et al.* Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.* **44**, 2846–2858 (2016).
  106. Abdelmohsen, K. *et al.* Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol.* **14**, 361–369 (2017).
  107. Lu, W. Y. Roles of the circular RNA circ-Foxo3 in breast cancer progression. *Cell Cycle* **16**, 589–590 (2017).
  108. Du, W. W. *et al.* Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death Differ.* **24**, 357–370 (2016).
  109. Yang, Q. *et al.* A circular RNA promotes tumorigenesis by inducing c-myc nuclear translocation. *Cell Death Differ.* (2017). doi:10.1038/cdd.2017.86
  110. Holdt, L. M. *et al.* Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat. Commun.* **7**, 12429 (2016).
  111. Chen, Y. G. *et al.* Sensing Self and Foreign Circular RNAs by Intron Identity. *Mol. Cell* **67**, 228–238.e5 (2017).
  112. Li, M. *et al.* Biogenesis of circular RNAs and their roles in cardiovascular development and pathology. *FEBS J.* (2017). doi:10.1111/febs.14191
  113. Kristensen, L. S., Hansen, T. B., Venø, M. T. & Kjems, J. Circular RNAs in cancer: opportunities and challenges in the field. *Oncogene* 1–11 (2017). doi:10.1038/onc.2017.361
  114. Reed, N. P., Henderson, M. A., Oltz, E. M. & Aune, T. M. Reciprocal regulation of Rag expression in thymocytes by the zinc-finger proteins, Zfp608 and Zfp609. *Genes Immun.* **14**, 7–12 (2013).

115. van den Berg, D. L. C. *et al.* Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron* **93**, 348–361 (2017).
116. Gruner, H., Cortés-López, M., Cooper, D. A., Bauer, M. & Miura, P. CircRNA accumulation in the aging mouse brain. *Sci. Rep.* **6**, 38907 (2016).
117. Vausort, M. *et al.* Myocardial Infarction-Associated Circular RNA Predicting Left Ventricular Dysfunction. *Journal of the American College of Cardiology* **68**, 1247–1248 (2016).
118. Peng, L. *et al.* Circular RNA ZNF609 functions as a competitive endogenous RNA to regulate AKT3 expression by sponging miR-150-5p in Hirschsprung's disease. *Oncotarget* **5**, 808–818 (2016).
119. Liu, C. *et al.* Silencing of circular RNA-ZNF609 ameliorates vascular endothelial dysfunction. *Theranostics* **7**, 2863–2877 (2017).
120. Wester, T. & Granström, A. L. Hirschsprung disease—Bowel function beyond childhood. *Semin. Pediatr. Surg.* **26**, 322–327 (2017).
121. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 1–9 (2016). doi:10.1136/gutjnl-2015-310912
122. Torre, L. A. *et al.* Global Cancer Statistics, 2012. *CA a cancer J. Clin.* **65**, 87–108 (2015).
123. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6**, 479–507 (2011).
124. Cancer Genom Atlas. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
125. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
126. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* **36**, 1461–1473 (2017).
127. Krausova, M. & Korinek, V. Wnt signaling in adult intestinal stem cells and cancer. *Cellular Signalling* **26**, 570–579 (2014).
128. Vatandoust, S., Price, T. J. & Karapetis, C. S. Colorectal cancer: Metastases to a single organ. *World J. Gastroenterol.* **21**, 11767–11776 (2015).
129. McIntyre, R. E., Buczacki, S. J. A., Arends, M. J. & Adams, D. J. Mouse models of colorectal cancer as preclinical models. *BioEssays* **37**, 909–920 (2015).
130. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
131. Etienne-Grimaldi, M. C. *et al.* K-Ras mutations and treatment outcome in colorectal cancer patients receiving exclusive fluoropyrimidine therapy. *Clin. Cancer Res.* **14**, 4830 (2008).



132. Santini, D. *et al.* High Concordance of KRAS Status Between Primary Colorectal Tumors and Related Metastatic Sites: Implications for Clinical Practice. *Oncologist* **13**, 1270–1275 (2008).
133. Oliveira, C. *et al.* KRAS and BRAF oncogenic mutations in MSS colorectal carcinoma progression. *Oncogene* **26**, 158–163 (2007).
134. Lee, S. Y. *et al.* Comparative genomic analysis of primary and synchronous metastatic colorectal cancers. *PLoS One* **9**, (2014).
135. Hosseini, H. *et al.* Early dissemination seeds metastasis in breast cancer. *Nature* **540**, 552–558 (2016).
136. Harper, K. L. *et al.* Mechanism of early dissemination and metastasis in Her2+ mammary cancer. *Nature* **540**, 588–592 (2016).
137. Zhao, Z.-M. *et al.* Early and multiple origins of metastatic lineages within primary tumors. *Proc. Natl. Acad. Sci.* **113**, 2140–2145 (2016).
138. Nguyen, D. X. Tracing the origins of metastasis. *Journal of Pathology* **223**, 195–204 (2011).
139. Klein, C. A. Cancer: The metastasis cascade. *Science* **321**, 1785–1787 (2008).
140. Liu, M. & Chen, H. The role of microRNAs in colorectal cancer. *J. Genet. Genomics* **37**, 347–58 (2010).
141. Han, D. *et al.* Long noncoding RNAs: Novel players in colorectal cancer. *Cancer Letters* **361**, 13–21 (2015).
142. Hackl, C. *et al.* Metronomic oral topotecan prolongs survival and reduces liver metastasis in improved preclinical orthotopic and adjuvant therapy colon cancer models. *Gut* **62**, 259–271 (2013).
143. Francia, G. *et al.* Long-term progression and therapeutic response of visceral metastatic disease non-invasively monitored in mouse urine using beta-human choriogonadotropin secreting tumor cell lines. *Mol. Cancer Ther.* **7**, 3452–3459 (2008).
144. Barrett, S. P. & Salzman, J. Circular RNAs: analysis, expression and potential functions. *Development* **143**, 1838–1847 (2016).
145. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nature Reviews Molecular Cell Biology* **18**, 637–650 (2017).
146. Sun, W. J. *et al.* RMBase: A resource for decoding the landscape of RNA modifications from high-throughput sequencing data. *Nucleic Acids Res.* **44**, D259–D265 (2016).
147. Müller, S. & Appel, B. In vitro circularization of RNA. *RNA Biology* **14**, 1018–1027 (2017).

148. Umekage, S. & Kikuchi, Y. In vitro and in vivo production and purification of circular RNA aptamer. *J. Biotechnol.* **139**, 265–272 (2009).
149. Ford, E. & Ares Jr, M. Synthesis of Circular RNA in Bacteria and Yeast using RNA Cyclase Ribozymes Derived from a Group I Intron of Phage T4. *PNAS* **91**, 3117–3121 (1994).
150. Chen, J. & Wagner, E. J. snRNA 3' end formation: the dawn of the Integrator complex. *Biochem. Soc. Trans.* **38**, 1082–7 (2010).
151. Yang, W., Du, W. W., Li, X., Yee, A. J. & Yang, B. B. 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 (2016).
152. Hinnebusch, A. G. The Scanning Mechanism of Eukaryotic Translation Initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
153. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* **11**, 113–127 (2010).
154. Jackson, R. J. The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
155. Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4**, 1–16 (2012).
156. Prabhakar, A., Capece, M. C., Petrov, A., Choi, J. & Puglisi, J. D. Post-termination Ribosome Intermediate Acts as the Gateway to Ribosome Recycling. *Cell Rep.* **20**, 161–172 (2017).
157. Petry, S., Weixlbaumer, A. & Ramakrishnan, V. The termination of translation. *Current Opinion in Structural Biology* **18**, 70–77 (2008).
158. Pisarev, A. V., Hellen, C. U. T. & Pestova, T. V. Recycling of Eukaryotic Posttermination Ribosomal Complexes. *Cell* **131**, 286–299 (2007).
159. Pisarev, A. V. *et al.* The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. *Mol. Cell* **37**, 196–210 (2010).
160. Becker, T. *et al.* Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* **482**, 501–506 (2012).
161. Young, D. J., Guydosh, N. R., Zhang, F., Hinnebusch, A. G. & Green, R. Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3'UTRs In Vivo. *Cell* **162**, 872–884 (2015).
162. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. Termination and post-termination events in eukaryotic translation. *Advances in Protein Chemistry and Structural Biology* **86**, 45–93 (2012).
163. Skabkin, M. A., Skabkina, O. V., Hellen, C. U. T. & Pestova, T. V. Reinitiation and other unconventional posttermination events during eukaryotic translation. *Mol. Cell*

- 51**, 249–264 (2013).
164. Belfort, M. *et al.* Processing of the intron-containing thymidylate synthase (td) gene of phage T4 is at the RNA level. *Cell* **41**, 375–382 (1985).
165. Chu, F. K., Maley, G. F., West, D. K., Belfort, M. & Maley, F. Characterization of the intron in the phage T4 thymidylate synthase gene and evidence for its self-excision from the primary transcript. *Cell* **45**, 157–166 (1986).
166. Puttaraju, M. & Been, M. Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Res.* **20**, 5357–5364 (1992).
167. Puttaraju, M. & Been, M. D. Circular ribozymes generated in Escherichia coli using group I self-splicing permuted intron-exon sequences. *J. Biol. Chem.* **271**, 26081–26087 (1996).
168. Baillat, D. & Wagner, E. J. Integrator: Surprisingly diverse functions in gene expression. *Trends in Biochemical Sciences* **40**, 257–264 (2015).
169. Baillat, D. *et al.* Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* **123**, 265–276 (2005).
170. Wang, K. *et al.* A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *Eur. Heart J.* **37**, 2602a–2611a (2016).
171. Liu, Q. *et al.* Circular RNA Related to the Chondrocyte ECM Regulates MMP13 Expression by Functioning as a MiR-136 ‘Sponge’ in Human Cartilage Degradation. *Sci Rep* **6**, 22572 (2016).
172. Xie, H. *et al.* Emerging roles of circRNA\_001569 targeting miR-145 in the proliferation and invasion of colorectal cancer. *Oncotarget* **7**, 26680–91 (2016).
173. Shang, X. *et al.* Comprehensive Circular RNA Profiling Reveals That hsa\_circ\_0005075, a New Circular RNA Biomarker, Is Involved in Hepatocellular Carcinoma Development. *Medicine (Baltimore)*. **95**, e3811 (2016).
174. Huang, M. *et al.* Comprehensive analysis of differentially expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in bladder carcinoma. *Oncotarget* **7**, 47186–47200 (2016).
175. Zhong, Z., Lv, M. & Chen, J. 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 (2016).
176. Nan, A. *et al.* A novel regulatory network among lncRpa, CircRar1, MiR-671 and apoptotic genes promotes lead-induced neuronal cell apoptosis. *Arch. Toxicol.* **91**, 1671–1684 (2017).
177. Chen, J. *et al.* Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. *Cancer Lett.* **388**, 208–219 (2017).
178. Xue, J. *et al.* Circ100284, via miR-217 regulation of EZH2, is involved in the

- arsenite-accelerated cell cycle of human keratinocytes in carcinogenesis. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1863**, 753–763 (2017).
179. Tang, C.-M. *et al.* {CircRNA}\_000203 enhances the expression of fibrosis-associated genes by derepressing targets of {miR}-26b-5p, {Col}1a2 and {CTGF}, in cardiac fibroblasts. *Sci. Rep.* **7**, 40342 (2017).
  180. Wang, K., Sun, Y., Tao, W., Fei, X. & Chang, C. Androgen receptor (AR) promotes clear cell renal cell carcinoma (ccRCC) migration and invasion via altering the circHIAT1/miR-195-5p/29a-3p/29c-3p/CDC42 signals. *Cancer Lett.* **394**, 1–12 (2017).
  181. Hsiao, K. Y. *et al.* Noncoding effects of circular RNA CCDC66 promote colon cancer growth and metastasis. *Cancer Res.* **77**, 2339–2350 (2017).
  182. Wu, Y., Zhang, Y., Zhang, Y. & Wang, J. J. CircRNA hsa\_circ\_0005105 upregulates NAMPT expression and promotes chondrocyte extracellular matrix degradation by sponging miR-26a. *Cell Biology International* (2017). doi:10.1002/cbin.10761
  183. Chen, L. *et al.* circRNA\_100290 plays a role in oral cancer by functioning as a sponge of the miR-29 family. *Oncogene* **36**, 4551–4561 (2017).
  184. Zhou, B. & Yu, J. W. A novel identified circular RNA, circRNA\_010567, promotes myocardial fibrosis via suppressing miR-141 by targeting TGF- $\beta$ 1. *Biochem. Biophys. Res. Commun.* **487**, 769–775 (2017).
  185. Han, D. *et al.* Circular RNA MTO1 acts as the sponge of miR-9 to suppress hepatocellular carcinoma progression. *Hepatology* (2017). doi:10.1002/hep.29270
  186. Dang, R. Y., Liu, F. L. & Li, Y. Circular RNA hsa\_circ\_0010729 regulates vascular endothelial cell proliferation and apoptosis by targeting the miR-186/HIF-1 $\alpha$  axis. *Biochem. Biophys. Res. Commun.* **490**, 104–110 (2017).
  187. Zhang, Y. *et al.* CircRNA\_100269 is downregulated in gastric cancer and suppresses tumor cell growth by targeting miR-630. *Aging (Albany. NY)*. **9**, 1585–1594 (2017).
  188. Zhong, Z. *et al.* Circular RNA MYLK as a competing endogenous RNA promotes bladder cancer progression through modulating VEGFA/VEGFR2 signaling pathway. *Cancer Lett.* **403**, 305–317 (2017).
  189. Wang, B.-G., Li, J.-S., Liu, Y.-F. & Xu, Q. MicroRNA-200b suppresses the invasion and migration of hepatocellular carcinoma by downregulating RhoA and circRNA\_000839. *Tumour Biol.* **39**, 1010428317719577 (2017).
  190. Zhang, X. L., Xu, L. L. & Wang, F. Hsa\_circ\_0020397 regulates colorectal cancer cell viability, apoptosis and invasion by promoting the expression of the miR-138 targets TERT and PD-L1. *Cell Biol. Int.* **41**, 1056–1064 (2017).
  191. Huang, X.-Y. *et al.* Comprehensive circular RNA profiling reveals the regulatory role of the circRNA-100338/miR-141-3p pathway in hepatitis B-related hepatocellular carcinoma. *Sci. Rep.* **7**, 5428 (2017).

192. Liang, H. F., Zhang, X. Z., Liu, B. G., Jia, G. T. & Li, W. L. Circular RNA circ-ABCB10 promotes breast cancer proliferation and progression through sponging miR-1271. *Am. J. Cancer Res.* **7**, 1566–1576 (2017).
193. Li, B. feng *et al.* Hsa\_circ\_0045714 regulates chondrocyte proliferation, apoptosis and extracellular matrix synthesis by promoting the expression of miR-193b target gene IGF1R. *Hum. Cell* **30**, 311–318 (2017).
194. Zhang, J. *et al.* Circular RNA\_LARP4 inhibits cell proliferation and invasion of gastric cancer by sponging miR-424-5p and regulating LATS1 expression. *Mol. Cancer* **16**, 151 (2017).
195. Tang, Y.-Y. *et al.* Circular RNA hsa\_circ\_0001982 Promotes Breast Cancer Cell Carcinogenesis Through Decreasing miR-143. *DNA Cell Biol.* (2017). doi:10.1089/dna.2017.3862
196. Pan, R.-Y. *et al.* Circular RNAs promote TRPM3 expression by inhibiting hsa-miR-130a-3p in coronary artery disease patients. *Oncotarget* **8**, 60280–60290 (2017).
197. Zhang, H. *et al.* Increased circular RNA UBAP2 acts as a sponge of miR-143 to promote osteosarcoma progression. *Oncotarget* (2017). doi:10.18632/oncotarget.18671
198. Guo, X.-Y. *et al.* circRNA\_0046367 Prevents Hepatotoxicity of Lipid Peroxidation: An Inhibitory Role against Hepatic Steatosis. *Oxid. Med. Cell. Longev.* **2017**, 1–16 (2017).
199. He, R. *et al.* circGFRA1 and GFRA1 act as ceRNAs in triple negative breast cancer by regulating miR-34a. *J. Exp. Clin. Cancer Res.* **36**, 145 (2017).
200. Deng, T. *et al.* Calcitonin gene - related peptide induces IL - 6 expression in RAW264.7 macrophages mediated by mmu\_circRNA\_007893. *Mol. Med. Rep.* (2017). doi:10.3892/mmr.2017.7779
201. Chen, J., Cui, L., Yuan, J., Zhang, Y. & Sang, H. Circular RNA WDR77 target FGF-2 to regulate vascular smooth muscle cells proliferation and migration by sponging miR-124. *Biochem. Biophys. Res. Commun.* **494**, 126–132 (2017).
202. Li, H. *et al.* CircFUT10 reduces proliferation and facilitates differentiation of myoblasts by sponging miR-133a. *J. Cell. Physiol.* (2017). doi:10.1002/jcp.26230
203. Wei, X. *et al.* Circular RNA profiling reveals an abundant circLMO7 that regulates myoblasts differentiation and survival by sponging miR-378a-3p. *Cell Death Dis.* **8**, e3153 (2017).
204. Yu, C. Y. *et al.* The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nat. Commun.* **8**, (2017).
205. Francia, G. *et al.* Comparative impact of trastuzumab and cyclophosphamide on HER-2-positive human breast cancer xenografts. *Clin. Cancer Res.* **15**, 6358–6366 (2009).

- 
206. Li, M. Z. & Elledge, S. J. SLIC: A method for sequence- and ligation-independent cloning. *Methods Mol. Biol.* **852**, 51–59 (2012).
  207. Cawthorne, C., Swindell, R., Stratford, I. J., Dive, C. & Welman, A. Comparison of doxycycline delivery methods for Tet-Inducible gene expression in a subcutaneous xenograft model. *J. Biomol. Tech.* **18**, 120–123 (2007).

## 8 CURRICULUM VITAE

### PERSONAL INFORMATION

First name / Surname: **Hung, HO XUAN**

Nationality: Vietnamese

Gender: Male

Date of birth: 06 July 1987

Permanent address: 62/6 Nguyen Khuyen, Nha Trang, Khanh Hoa, Vietnam

Permanent e-mail: [hunghok9@gmail.com](mailto:hunghok9@gmail.com)

### EDUCATION

2014 - 2018: PhD in Biochemistry, University of Regensburg, Germany

2011 - 2013: Master of Science in Biology, Major Biochemistry, ETH Zurich Switzerland

2005 - 2009: Bachelor Degree of Talented program in Biology, Vietnam National University, Hanoi

### TRAINING

March 2014 – 2018: Biochemical characterization of non-coding RNPs. Supervisor: Prof. Dr. Gunter Meister

January 2012 – February 2013: Mechanisms of influenza A virus (IAV) infection. Supervisor: Prof. Ari Helenius, Dr. Yohei Yamauchi (Master Thesis)

2007 – 2009: Training projects in Key Laboratory of Enzymes and Protein Technology, VNU, Hanoi, Vietnam

### HONORS AND AWARDS

March 2014 – February 2017: Marie Curie Fellowship, RNAttrain network, Early Stage Research fellow

March 2013 – September 2013: Marie Curie Fellowship, virus entry network, internship

2011 – 2013: Swiss Government Excellence Scholarship

June 2009: Certificate of merit excellent study achievements during Bachelor studies (2005-2009), Vietnam National University, Hanoi

September 2005: Certificate of merit for First Place in the National University Entrance Exam for high-school students, Vietnam National University, Hanoi

**PROFESSIONAL EXPERIENCE****Research Assistant, March 2013 – August 2013**

- Employer: Institute of Veterinary Medicine, Free University Berlin, Philippstr. 13, Berlin, Germany
- Main responsibilities: Participating in the project Molecular Mechanism of Cell entry of Enveloped Viruses belonged to the 'Virus Entry' Networks of the Marie Curie Program FP7-People-ITN-2008

**Research Assistant, June 2010 – July 2011**

- Employer: National Institute of Hygiene and Epidemiology, 01 Yersin, Hanoi, Vietnam
- Main responsibilities: Being involved in 2 projects in immunology and epidemiology
  1. Seroprevalence of Avian Influenza in Vietnam, funded by and under auspices of National Institutes of Health (USA), 2010-2011
  2. Identification of mechanisms correlating with susceptibility for Avian Influenza (IMECS), funded by European Community; 2009-2011

**LABORATORY SKILLS**

- qPCR and molecular cloning
- Protein expression, extraction and purification
- Tissue culture, CRISPR/Cas9, Fluorescent microscopy
- Northern blotting, Western blotting, Co-immunoprecipitation

**TEACHING SKILLS**

- Bachelor thesis (Lisa Seider)
- Practical course (Nicole Kabella, Simon Eder, Kevin Heizler)
- Masters thesis (Annamaria Dalfino-Spinelli, Kevin Heizler)



## 9 Publications and Presentations

### 9.1 Publications

1. **Hung Ho-Xuan**, Petar Glazar, Marvin Anders, Claudia Latini, Kevin Heizler, Ivano Legnini, Gerhard Lehmann, Norbert Eichner, Irene Bozzoni, Nikolaus Rajewsky, Christina Hackl and Gunter Meister. Circular RNA ZNF609 promotes the progression of colorectal cancer. (Manuscript in preparation)
2. Nguyen Thi Hong Loan, **Ho Xuan Hung**, Ngo Thi Huyen Trang, Nguyen Thi Van Anh, Phan Tuan Nghia, (2010), "Cloning and expression of protease-encoding gene of HIV type 1 isolated from Vietnam", Vietnamese Journal of Biotechnology, 8(2).pp. 227-233

### 9.2 Presentations

1. **Hung Ho-Xuan**, Petar Glazar, Marvin Anders, Claudia Latini, Ivano Legnini, Gerhard Lehmann, Norbert Eichner, Irene Bozzoni, Nikolaus Rajewsky, Christina Hackl and Gunter Meister. Circular RNA ZNF609 promotes the progression of colorectal cancer. **Poster** presented at SFB 960 Symposium: The Biology of RNA-Protein Complexes, Regensburg, 2017
2. **Hung Ho-Xuan**, Petar Glazar, Marvin Anders, Claudia Latini, Ivano Legnini, Gerhard Lehmann, Norbert Eichner, Irene Bozzoni, Nikolaus Rajewsky, Christina Hackl and Gunter Meister. Circular RNA ZNF609 promotes the progression of colorectal cancer. **Oral presentation** presented at EMBL/EMBO Symposium: The non-coding genome, Heidelberg, 2017
3. **Hung Ho-Xuan**, Petar Glazar, Marvin Anders, Claudia Latini, Ivano Legnini, Gerhard Lehmann, Norbert Eichner, Irene Bozzoni, Nikolaus Rajewsky, Christina Hackl and Gunter Meister. Circular RNA ZNF609 promotes the progression of colorectal cancer. **Poster** presented at RNA society meeting, Prague, 2017



## 10 Acknowledgements

My first word of gratitude must go towards Gunter Meister, whose kindness, knowledge and enthusiasm have inspired me during the whole time in the lab as well as in future career. My PhD project fits perfectly to my expectation of being a PhD student: freedom in research.

My second appreciation is reserved for Matthias Stotz for his great academic encouragement, his patience, and support during my first 6 months in Meister group.

I am very thankful:

...to Anders Lund and Markus Kretz for being my PhD thesis mentors and giving me helpful suggestion for project,

...to Christoph Klein, Markus Kretz, Wolfgang Seufert and Dina Grohmann for being in my PhD thesis examination committee,

...to all people who contributed to this work. Firstly, the project has constantly been supported by Christina Hackl and Marvin Anders for the mouse work and discussion. Secondly, I thank Petar Glazar, Foivos Gypas, Nikolaus Rajewsky and Mihaela Zavolan for bioinformatics analysis; Norbert Eichner for miRNA library preparation and data analysis; Gerhard Lehmann for differential gene expression analysis; Debbie van den Berg for providing anti-Zfp609 antibodies; Rüdiger Eder and Irina Fink for cell sorting with FACSaria IIu; Astrid Bruckmann and Eduard Hochmuth for mass spectrometry analysis; So Umekage and Irene Bozzoni for providing plasmids; Hedayatollah Hosseini and Christoph Klein for lentiviral experiments in S2 lab. Last but not least, Claudia Latini, Kevin Heizler, Daniele Hasler and Franziska Weichmann for troubleshooting different parts of the project; Sigrun Ammon and Corinna Friederich for preparing reagents and buffers.

...to the whole research group and Biochemistry I Institute for being very helpful and providing a very nice working atmosphere. Special thanks go to Johannes Danner, Claudia Latini, Balagopal Pai for helping me with PhD thesis submission. I am also grateful for fruitful discussion with the former colleagues of the lab: Anne Dueck, Leonhard Jakob, and Melina Musri.

...to all friends and colleagues of the Marie Curie RNAttrain network; all friends in Vietnam; Anna Kuhn and Christian Reimelt for a nice time in Regensburg; Johannes Danner and Christian Ziegler for being good friends in the lab and having nice time after work.

... to European Union for funding via FP7 Marie Curie scheme.

Cám ơn bố mẹ, đại gia đình hai bác, gia đình anh Mạnh, em Dung, gia đình em Lô đã luôn sát cánh trong những năm tháng ở trời Âu.

This thesis is dedicated to my family, especially to my parents, Minh Minh, Lô and Moon.



*This work is ...  
to be continued....*

