Chapter Five – Development of reporter systems for probing the miRNA pathway

1. Introduction

1.1. Why study miRNA biogenesis and effector pathways?

Understanding miRNA biogenesis and downstream effector pathways is important to reveal crucial regulation points which intersect with other molecular pathways to control phenotypic outcomes in response to external stimuli in different physiological and pathological environments.

Recent advances have brought significant understanding in the field of miRNA biogenesis and downstream effector functions. Many of the components in the miRNA biogenesis pathway have been identified from siRNA-mediated gene silencing and biochemical studies such as pulling down protein complexes associated with known components of the system (Bernstein *et al.*, 2001; Hutvagner *et al.*, 2001; Lee *et al.*, 2003; Han *et al.*, 2004). Although much has been learnt, there are likely to be novel enzymes and regulators yet to be discovered. There are many interesting but outstanding questions regarding mammalian miRNA processing and effector pathways. Questions such as, what are the determinants in specifying the two downstream effector pathways, miRNA-mediated mRNA cleavage or post-translational inhibition? How are RISC complexes containing mature miRNAs degraded or recycled? Is there an "amplifying system" in mammals, equivalent to the system observed in worms and plants, to generate secondary siRNAs to efficiently mediate gene repression?

Forward genetic screens have been conducted in the invertebrate model organisms *C. elegans* and *Drosophila* to discover genes in siRNA-mediated silencing pathway (Kim *et al.*, 2005; Dorner *et al.*, 2006). Dorner and co-workers used dsRNA-mediated RNAi to knockdown genes on a genome-wide scale in cultured Drosophila S2R+ cells and followed by transfection of dsRNAs targeting the luciferase reporter system as readout for detecting RNAi pathway abrogation. This dsRNA-based RNAi screening system has identified seven genes, including a known gene *Ago2* and two members of the heat shock protein family members *Hsc70-4 and*

Hsp70-3 (Dorner et al., 2006). The specificity and efficacy of the dsRNA-mediated gene knockdown may be limited in this system and the precise function in the RNAi pathway of the novel genes found is yet to be discovered. In the *in vivo C. elegans* study, Kim *et al* engineered a GFP reporter and a dsRNA hairpin targeting GFP in the epithelial seam cells in *C. elegans* to give rise to the "RNAi sensor" strain (Kim et al., 2005). An RNAi screen was conducted by feeding the RNAi sensor strain with specific dsRNAs to knockdown 19,000 genes in the worm genome. Mutant worms with the GFP expression restored in the seam cells may implicate the function of targeted genes in the RNAi pathway. In this screen, 90 genes were identified to have RNAi mutant phenotype and 11 of which are known RNAi components. The newly identified genes include members of the nonsense-mediated decay (NMD) pathway, members of the pre-mRNA cleavage and polyadenylation complex, and factors involved in nuclear transport and chromatin factors. Several genes found in this screen are relevant to the miRNA pathway as miRNA and siRNA share similarities and are artificially interchangeable. However, *Kim* et al found that several factors are RNAi specific and had little molecular overlap to the miRNA pathway by examing the *let-7*-associated supernumerary seam cell phenotype (Kim et al., 2005). This work suggests that miRNA pathway possess unique factors that are not shared with the RNAi pathway. In addition, mammalian systems might use and regulate these pathways differently to invertebrates. Genetic screens performed in mammalian systems may be able to discover mammalian specific factors. Despite the advantages of non-hypothesis driven approaches in mammalian systems to identify components of these pathways, so far they have not been conducted.

Finally, there is a strong clinical motivation to investigate miRNA biogenesis and effector pathways. It is widely accepted that there is a tight connection between the miRNA pathway and cancer (Lu et al., 2005), which were described in detail in Chapter One. Mutations in miRNA processing genes have been found in a wide range of human cancers and some well-known tumour suppressors, such as p53, have been reported to be directly involved in the biogenesis of several miRNAs in DNA damage response (Suzuki et al., 2009) (Chapter One). Therefore, novel components in miRNA biogenesis and effector pathways may provide insights into the function of genes mutated in cancer whose functions are yet to be defined.

1.2. Challenges in using *Blm*-deficient ES cell system to study the miRNA pathway

One of the main constraints in employing current mixed population-based recessive genetic screens in using the *Blm*-deficient ES cell system is the extremely low ratio between the relevant homozygous mutants to irrelevant cells within the pool. Therefore, a potent but sensitive selection/isolation method is required in order to identify the few genuine homozygous mutants out of approximately 30 million irrelevant cells. This poses a major challenge when one is using the *Blm*-deficient ES cell system to investigate biological pathways that are not directly selectable. Therefore, screen designs, which are able to "translate" the characteristics of a mutant phenotype in the biological pathway of interest into a strong reporter system, is required to conduct the screen.

The miRNA processing and effector pathway falls into a category of "subtle" biological pathways in which mutant phenotypes can not be directly assessed in mouse ES cells. Therefore, a sensitive reporter system is pivotal for the success of the screen. This chapter describes the development of two reporter systems to probe the miRNA pathway, with each designed to specifically explore a branch of the pathway. Both reporter systems share common elements in miRNA biogenesis, thus mutants which affect biogenesis can be identified using both systems.

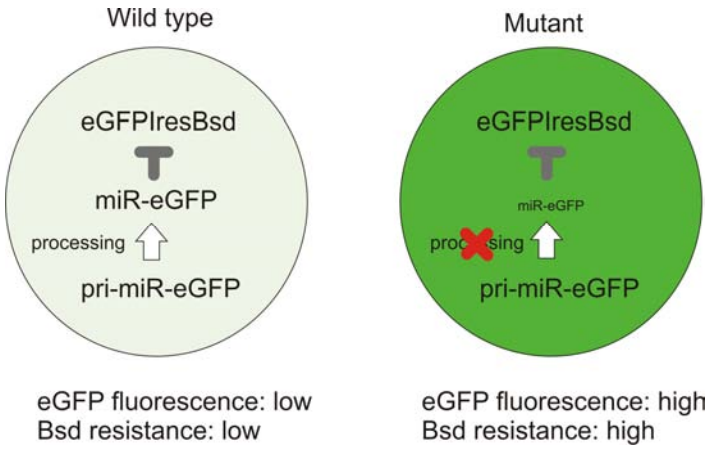
2. Results

2.1. Development of the artificial miR-eGFP system

The first reporter system I developed is based on an artificial miRNA with a seed sequence that perfectly matches the *eGFP* coding region to achieve miRNA-mediated *eGFP* mRNA cleavage. In miRNA pathway proficient cells, *eGFP* mRNA is expressed but cleaved and these cells do not express eGFP protein. A change in fluorescence intensity from non-green to green is a direct phenotypic readout of mutants which affect miRNA processing and effector pathways. Due to the extremely low ratio of mutant to background cells, a drug selection system was also incorporated into the reporter target. To achieve this, the *eGFP* coding region is linked to the Blasticidin S deaminase gene (Bsd) via a viral *IRES* sequence, so that *eGFP* and *Bsd* are transcribed as a single mRNA. The repression of *eGFP* caused by artificial miR-eGFP-

mediated mRNA cleavage can also prevent translation of Bsd, hence a miRNA biogenesis mutant will be more resistant to blasticidin selection than corresponding miRNA-pathway proficient cells. Thus blasticidin resistance can be used to directly select homozygous mutants from a mixed pool. Background mutants which affect Bsd resistance by other mechanisms can be excluded by further examination of the eGFP expression level. Figure 5-1 illustrates this reporter system.

Figure 5-1: Schematic representation of the artificial miR-eGFP system.



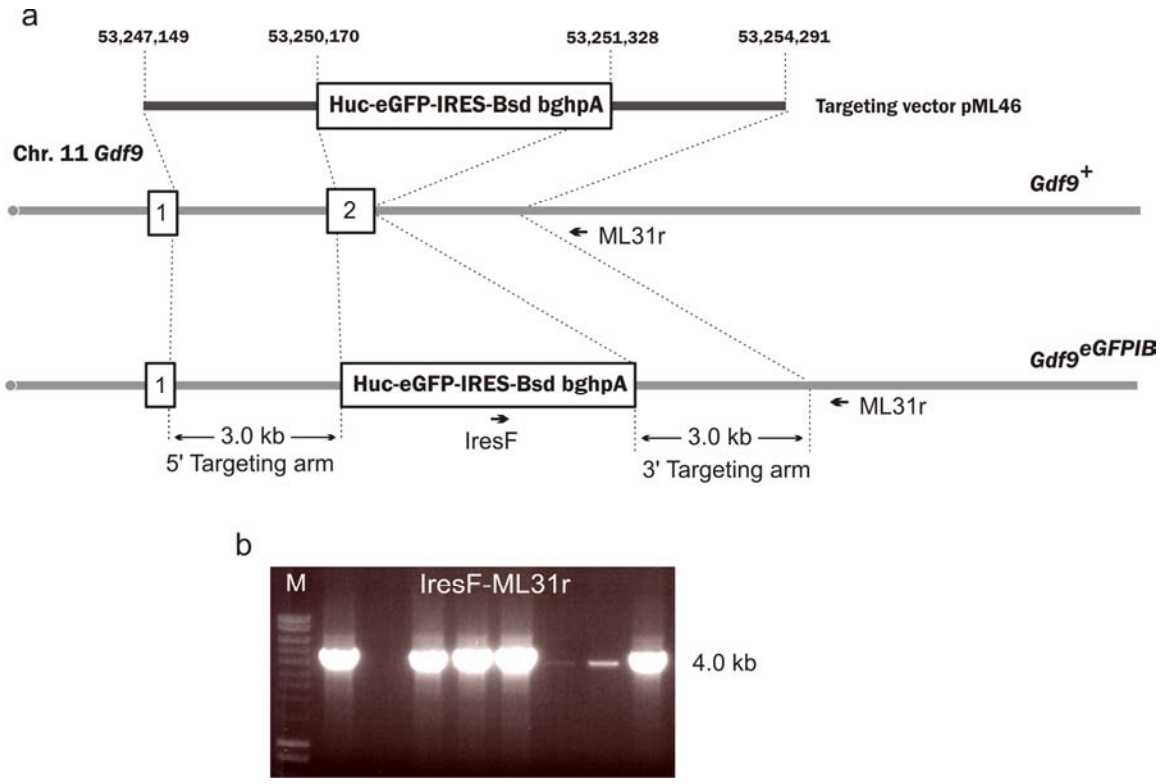
In this reporter design, the essential components are the eGFP-IRES-Bsd reporter and a miR-eGFP which efficiently represses its expression. This system needs to be stable yet sensitive enough to distinguish processing mutants from non-mutant cells. The eGFP-IRES-Bsd construct has been described in Chapter 4. It was targeted into both alleles of the *Blm* locus while making *Blm*-deficient ES cells (Chapter 4). The homozygosity of this reporter is an important design feature of this screen since in a heterozygous context, a major background in the screen would have been the loss of the heterozygosity of this reporter.

An artificial miR-eGFP system was generated based on human miR-30 and mouse miR-155 backbones to generate miRNAs with target recognition sequences to the eGFP coding region. The target recognition sequences were designed to be perfectly complementary to the eGFP sequence, thus the effector pathway downstream is expected to be miRNA-mediated mRNA cleavage.

2.1.1. Generation of an eGFP-knockin cell line

While the *Blm*-deficient ES cell line was under construction, an eGFP-IRES-Bsd knockin cell line was generated in order to validate the artificial miR-eGFP designs using the *Blm*-deficient NN5 ES cells. A targeting vector (pML46) was constructed to introduce the eGFP-IRES-Bsd reporter into the *Gdf9* locus, giving rise to a new cell line NN5- Gdf9$^{eGFP/+}$. The targeting vector pMl46 was linearised by *Not*I and electroporated into 1x10^{7} NN5 cells and plated into a 90 mm culture plate with a mono-layer of G418 and blasticidin double-resistant feeders. The following day, blasticidin selection (10 µg/ml) was initiated. The cells were selected for seven days and 48 colonies were picked and genotyped using long-range PCR (with primers IresF and ML31r) to identify homologous recombination events at the 3' targeting arm, Figure 5-2.

Figure 5-2: Targeting *Gdf9* locus with an eGFPIRESBsd reporter.

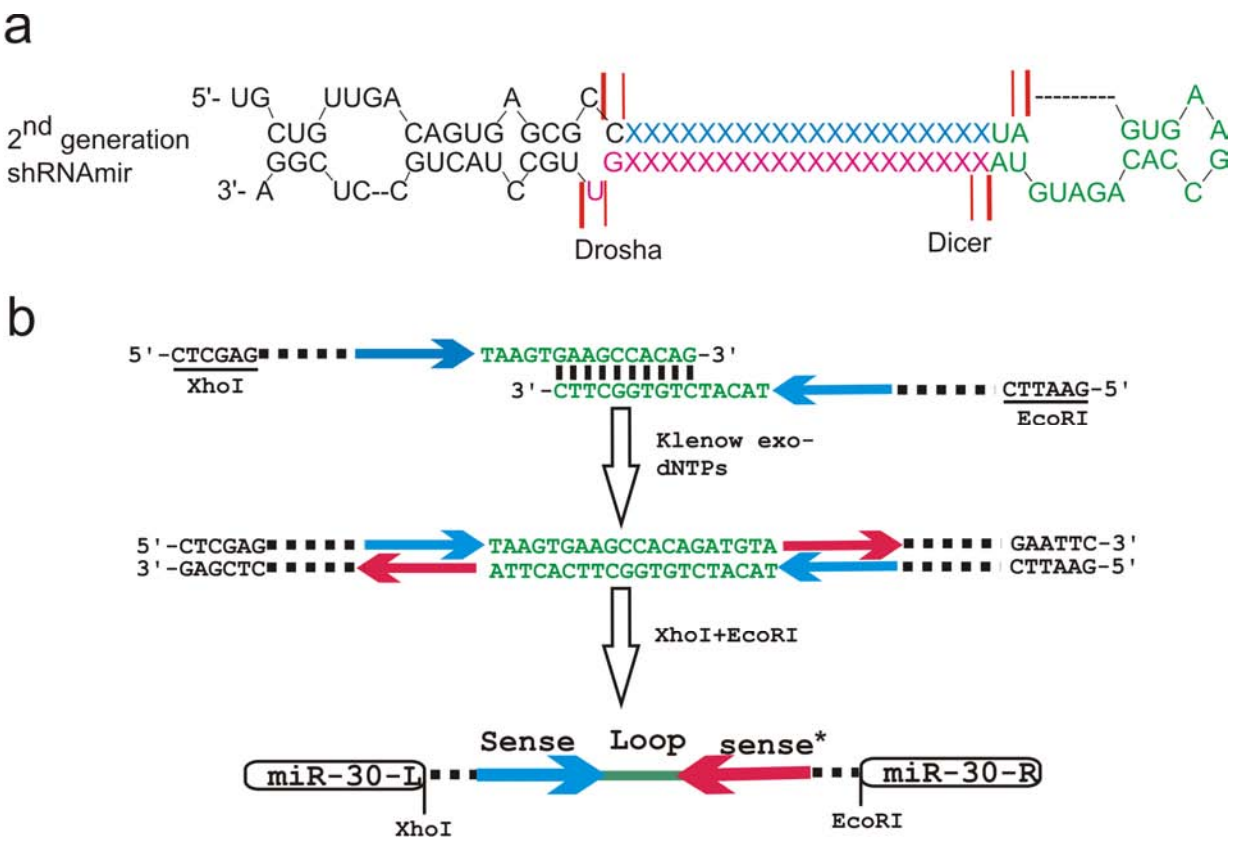


a, Targeting strategy of the *Gdf9* locus with the reporter construct b, Long-range PCR confirmation of positively targeted ES clones using primers Ires-F and ML31r. The genomic fragments used as targeting arms are indicated in the targeting vector with coordinates based on NCBI build37.

2.1.2. miR-30-based miR-eGFP generation and analysis

The miR-30-based shRNA (Figure 5-3a) has been shown to mediate efficient gene knockdown previously (Silva et al., 2005). A schematic representation of the procedures used to generate customised miR-eGFP using the human miR-30 backbone is shown in Figure 5-3b. A 20 nt eGFP recognition sequence and its complementary sequence were incorporated into two synthetic oligo-nucleotides juxtaposed to the miR-30 loop region. The two oligos were annealed and subsequently extended by Klenow exo$^-$ fragment to give rise to a double strand DNA fragment coding for the miR-eGFP. The miR-eGFP fragment was then digested with restriction enzymes and cloned into a plasmid. The miR-eGFP backbone and its target-sequence region were confirmed by sequencing. Two different miR-eGFP target recognition sequences used in this experiment were designed based on experimentally-validated eGFP knockdown from published literature. These two target sequences recognise different parts of the eGFP coding region and were used to construct miR-30-based miR-eGFP, and they were named no.44 and no.91.

Figure 5-3: The construction of the miR-30 backbone based single-unit miR-eGFP.

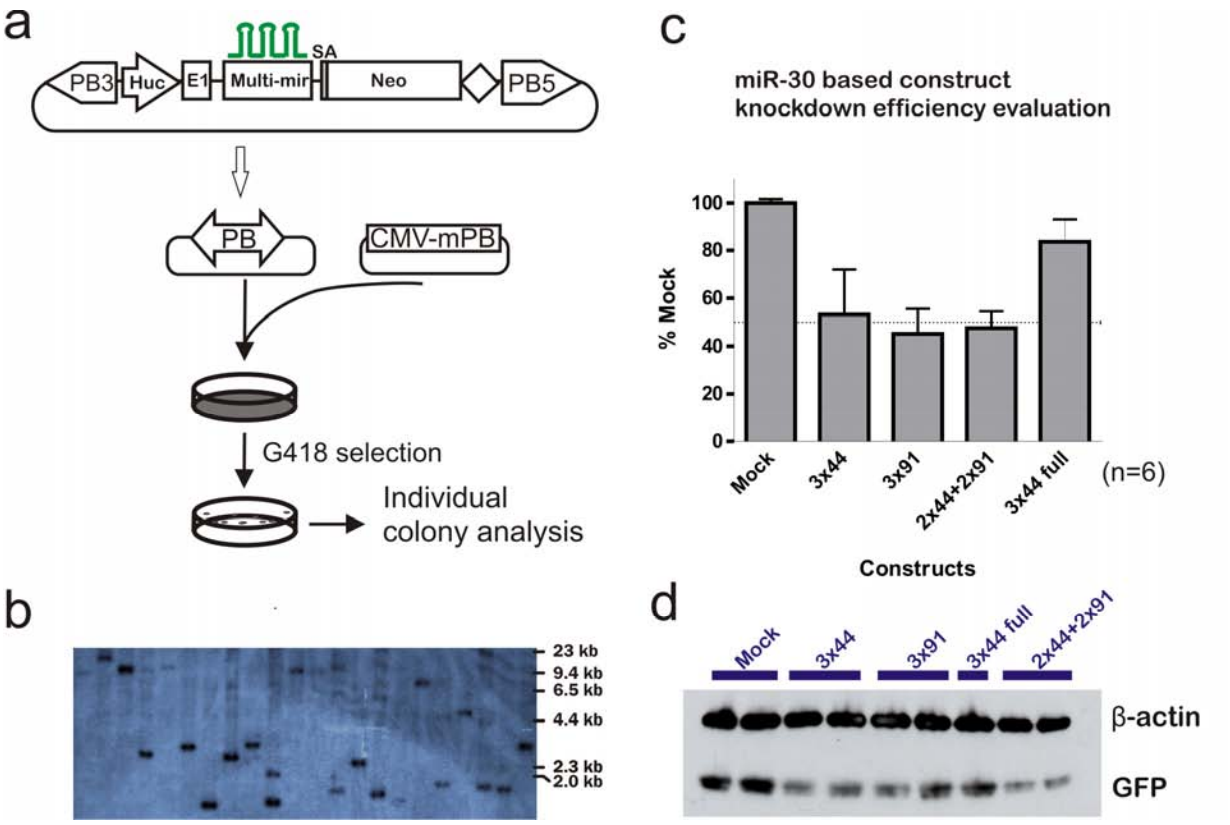


a, The human miR-30 backbone used for engineering of second-generation shRNAmir described by Silva and co-workers, and the diagram is adapted from (Silva et al., 2005). The 3' strand (pink) is the mature miRNA post processing. The red lines represent the Drosha and Dicer processing sites while the heavy and thin red lines represent the major and minor processing sites respectively. b, The cloning method used to generate the miR-30 based miR-eGFP. Two long synthetic oligos including the 20 nt eGFP targeting sequencing (the blue and red arrows which represent the sense and anti-sense strand sequences respectively) were annealed with ten nucleotides overlapping sequence at the loop region. The partially annealed oligos were treated with Klenow fragment (exo^-) to form a dsDNA fragment. The fragment was then cut with restriction enzyme at the sites indicated and cloned into a plasmid with the corresponding restriction enzyme sites. The colour scheme in a and b is correlated.

In order to increase the repression efficiency, multiple copies of the miR-eGFP was constructed in a head-to-tail fashion to give rise to miR-eGFP multimers "3x44" (three copies of miR-eGFP with no.44 as the eGFP recognition sequence), "3x91" (three copies of miR-eGFP with no.91 as the eGFP recognition sequence) and "2x44+2x91" (two copies of miR-eGFP with no.44 and no.91 as the eGFP recognition sequence each). Finally, a miR-eGFP multimer "3x44 full" was also generated with three copies of the miR-eGFP with no. 44 as the recognition sequence. However in this case, each miR-eGFP includes 200 bp of the DNA sequences flanking the miR-30 hairpin, as it has been proposed that this "context" can enhance the miRNA initial processing by the Drosha complex (Silva et al., 2005). The miR-eGFP multimers were inserted into intron 1 of the human unbiquitin C (huc) fragment containing the promoter region to generate a miR-eGFP and neomycin resistant gene co-expressing construct, Figure 5-4a. This architecture ensures the constant expression of miR-eGFP in G418 resistant ES cells when the construct is stably integrated. The construct was further introduced within the PB transposon, allowing the stable and simple delivery of the constructs to reporter ES cells, Figure 5-4a. Using the PB transposon and PBase co-electroporation system, stable integration of the PB transposon can be achieved with one copy per cell by electroporating $1x10^7$ NN5-Gdf9$^{eGFP/+}$ cells with 100 ng of the PB transposon plasmid and 5 μg of the mPBaseΔNeo plasmid. Southern blotting, shown in Figure 5-4b, confirmed that the majority of the G418 resistant colonies generated under these conditions contain a single copy of the PB transposon.

Plasmids with the PB transposons carrying different designs and combinations of the miR-eGFP multimers were co-electroporated with mPBase into the NN5-Gdf9$^{eGFP/+}$ reporter cell line. After eight days of G418 selection, six colonies were picked from each transfection corresponding to the different miR-eGFP designs and these clones were analysed for eGFP expression, Figure 5-4c. As a control, NN5-Gdf9$^{eGFP/+}$ cells were transfected with the PB transposon containing the Huc-Neo cassette without the miR-eGFP (mock).

Figure 5-4: miR-30-based artificial miR-eGFP evaluation.



a, Three copies of the miR-30-based miR-eGFP were incorporated within the intron of a neomycin resistant cassette (Neo) cassette driven by the human ubiqiutin C (Huc) promoter. The exon 1 and the intron were derived from the endogenous human ubiquitin C gene and the splice acceptor from the endogenous exon 2 of the human ubiquitin C gene was placed upstream of the *Neo* coding region. The entire expression cassette was cloned into the PB transposon. The miR-eGFP expression construct was delivered into the NN5- Gdf9$^{eGFP/+}$ using the PB transposon system under single copy per cell delivery conditions. b, Southern blot using PB5'ITR as the probe to determine the copy number of the PB-miR-eGFP-Neo integrations. The genomic DNA was digested with *Pst*I. c, flow cytometry analysis of eGFP intensity in G418 resistant colonies. Six colonies were analysed for each type of construct. d, western blot confirmation of the eGFP protein level in G418 resistant colonies.

All miR-eGFP designs achieved a degree of eGFP knockdown. Both eGFP target recognition sequences ("3x44" and "3x91") performed similarly, giving approximately 50 % of the eGFP repression. The combination of the two different eGFP target recognition sequences

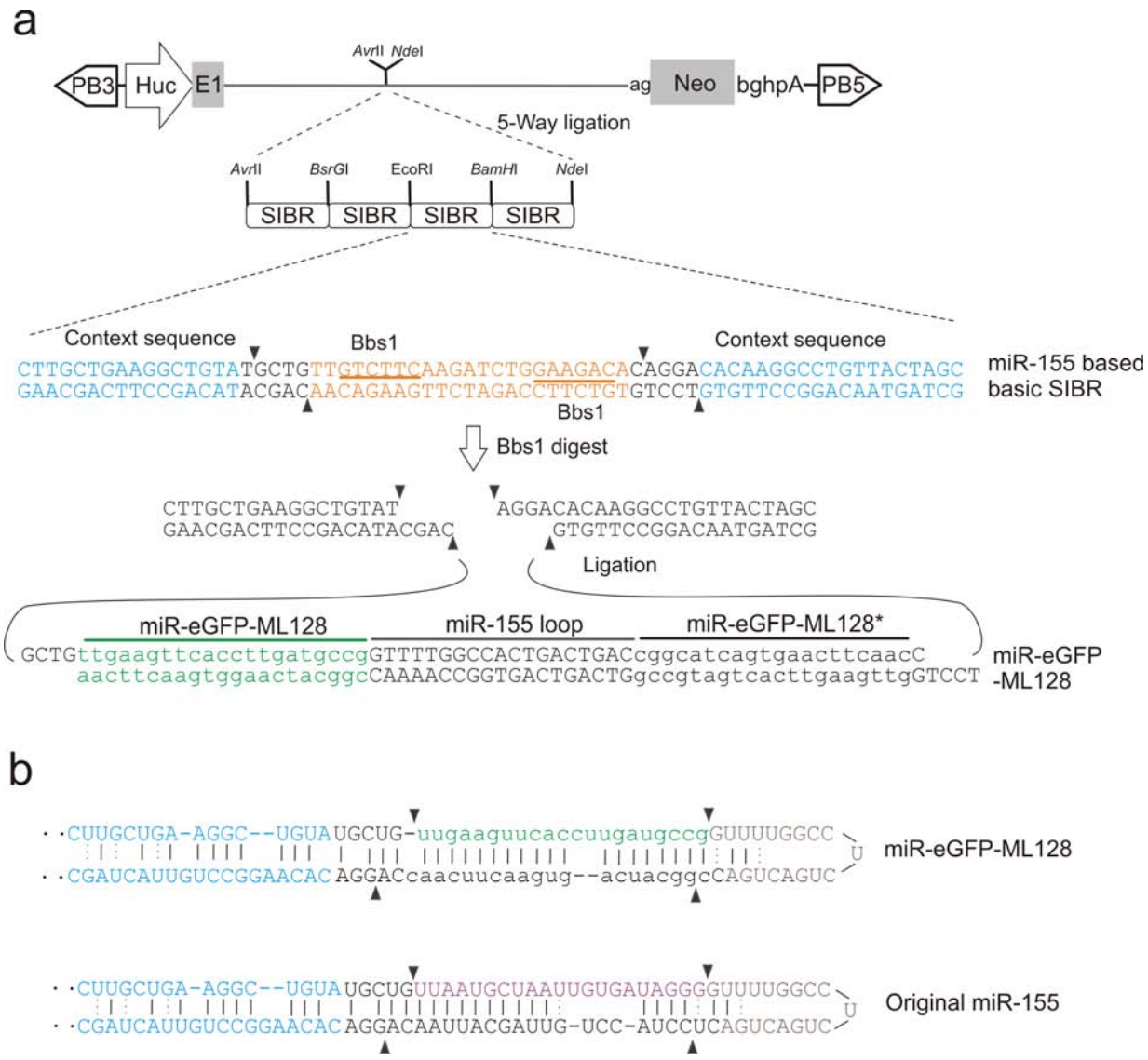
("2x44+2x91") did not result in synergistic repression, the level of repression was comparable to individual eGFP recognition sequences of similar copy numbers. However, the miR-eGFP multimer, which included the miR-30 flanking sequences ("3x44 full"), performed worse than the smaller version ("3x44"). This may reflect synergistic processing of miR-eGFP multimers, as in the "3x44 full" construct, the individual miR-eGFP hairpins were separated by 400 bp of sequence. When the miR-eGFP hairpins are directly next to each other, they might facilitate the recognition of Drosha on adjacent processing sites or they might enhance pri-miRNA processing itself. eGFP protein levels of representative clones were further analysed from all types of the miR-eGFP constructs by Western blotting. The eGFP protein levels showed a similar level of repression to the results obtained from fluorescent analysis, Figure 5-4d.

2.1.3. Construct optimisation for the artificial miR-eGFP construct

The observation of repression in eGFP suggests that the artificial miR-eGFP generated can be processed to mediate an eGFP-specific "knockdown". However, the level of repression achieved by all constructs was not sufficient for screening purposes. The inefficiency of the knockdown could be due to inefficient processing of miR-30, the target recognition sequence itself not being sufficient in mediating repression or the promoter used is not being strong enough to drive adequate levels of the miR-eGFP.

To optimise the knockdown efficiency, a set of constructs were generated using different promoters, target sequences and miRNA backbones. A panel of total 12 constructs (Table 5-1) were tested, including three different promoters (Huc, CAG and human EF1α promoters), five different target sequences (44, 91, 126, 127, and 128, the actual sequences are described in the methods in Chapter 2), two miRNA backbones (mouse miR-155 and human miR-30) and various copy numbers of the multimer. The mouse miR-155 backbone was based on the design of synthetic inhibitory BIC-derived RNA (SIBR) cassette, constructed by Chung and co-workers (Chung et al., 2006). The precise sequence of the SIBR cassette and structure of the miR-155 backbone based miR-eGFP multimer is shown in Figure 5-5 with miR-eGFP 128 as an example. The constructs were evaluated for the eGFP knockdown efficiency as described for the miR-30 backbone (Figure 5-4a).

Figure 5-5: miR-155 backbone based miR-eGFP multimer generation.



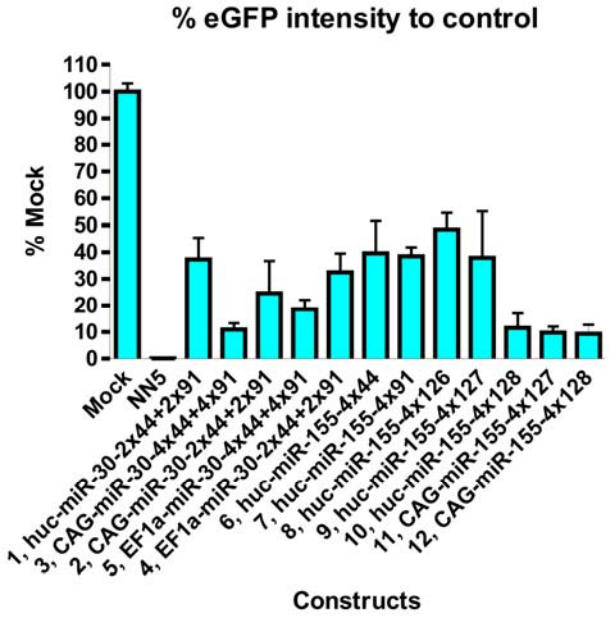
a, The design of the miR-eGFP and Neo co-expression cassette in PB transposon. Each synthetic inhibitory BIC-derived RNA (SIBR) cassette contains a single miR-eGFP unit with the miR-155 as the backbone. The eGFP recognition target and its antisense sequence containing oligos were annealed and the double-strand fragment was ligated into *Bbs*I digested SIBR cassette. b, An example of a miR-eGFP (with target sequence 128) compared to the original miR-155 hairpin structure. The wobble base pairing is shown with dotted lines, whereas the full base pairing is illustrated with straight lines. The black triangles indicate the Drosha and Dicer main processing sites.

Table 5-1: Summary of different miR-eGFP constructs generated.

Basic structure

Construct	Promoter	Backbone	Target seq.	Copy No.
1	Huc	miR-30	44 and 91	2 each
2	CAG	miR-30	44 and 91	2 each
3	CAG	miR-30	44 and 91	4 each
4	EF1a	miR-30	44 and 91	2 each
5	EF1a	miR-30	44 and 91	4 each
6	Huc	miR-155	44	4
7	Huc	miR-155	91	4
8	Huc	miR-155	126	4
9	Huc	miR-155	127	4
10	Huc	miR-155	128	4
11	CAG	miR-155	127	4
12	CAG	miR-155	128	4

Figure 5-6: Optimisation of an efficient miR-eGFP to knockdown eGFP.



Analysis of eGFP intensity for each construct. NN5, the parental cell line without the eGFP reporter knock-in. Mock, transfection of a plasmid containing the PB transposon with the PGK-Neo cassette. The values shown are an average of four clones assessed for each construct.

The performance of each miR-eGFP construct was assessed by electroporation of $1x10^7$ NN5-Gdf9$^{eGFP/+}$ cells with 100 ng of each construct together with 5 µg of the CMV-mPBΔNeo plasmid. The electroporated cells were plated on a 90 mm plate and selected with G418. After eight days selection, four colonies were picked from each plate and the colonies were expanded and were subjected to fluorescence analysis. The results of eGFP intensity measurements of all constructs are shown in Figure 5-6.

Comparing the eGFP repression among constructs 6-10, which have identical promoter and miR-eGFP copies but have different target sequences, target sequence 128 showed 90 % repression of the control eGFP level, whereas other sequences only resulted in 50 % - 60 % repression. Thus, the nature of the sequence is the major determinant of the knockdown efficiency.

The effect of different promoters can be compared with constructs 1, 2 and 4, which contain identical target sequences. The CAG-driven miR-eGFP multimer gave the best knockdown efficiency as expected, since CAG is the strongest promoter of the three and Huc is the weakest. High levels of expression improve the degree of knockdown by providing the miRNA processing reactions with more substrate. However, the effect is reduced when the knockdown is already relatively efficient as shown by comparing constructs 10 and 12. This may reflect the fact that in cases where the knockdown is efficient, the miRNA processing and effector silencing machineries are close to saturation.

The effect of copy number of the miR-eGFP within the multimer can also be compared with constructs 2 and 3 as well as constructs 4 and 5. Increasing numbers of hairpins provided only a moderate increase in the knockdown efficiency. The synergistic effect observed previously (Figure 5-4c, “3x44 full” *vs* “3x44”) is most dramatic when the copy number increases from one to three (data not shown). When additional copies were incorporated, the synergistic effect diminished. Thus the slight increase in efficiency is due to the increase in precursor miR-eGFP substrate only.

Finally, the miRNA backbone can be compared in construct 1 *vs* 6 and 7. It was found from the previous experiment (Figure 5-4c) that target sequences 44 and 99 perform similarly in eGFP knockdown. Thus construct 1 with the combination of two different target sequences can be compared to constructs 6 and 7 with target sequence 44 and 99 singly, respectively. miR-30 and miR-155 backbone did not show any difference in the repression efficiency.

From these comparisons, miR-eGFP with sequence 128 in a multimer format provided the best repression. Although the CAG promoter was optimal in this analysis, it may be prone to silencing in mouse ES cells due to the presence of cytomegalovirus enhancer sequence within the promoter. Since the promoter strength is less significant when an optimal target sequence was used, Huc driven miR-eGFP multimer with sequence 128 was selected for my reporter miR-eGFP construct as this will provide a comparable repression to the CAG-driven construct.

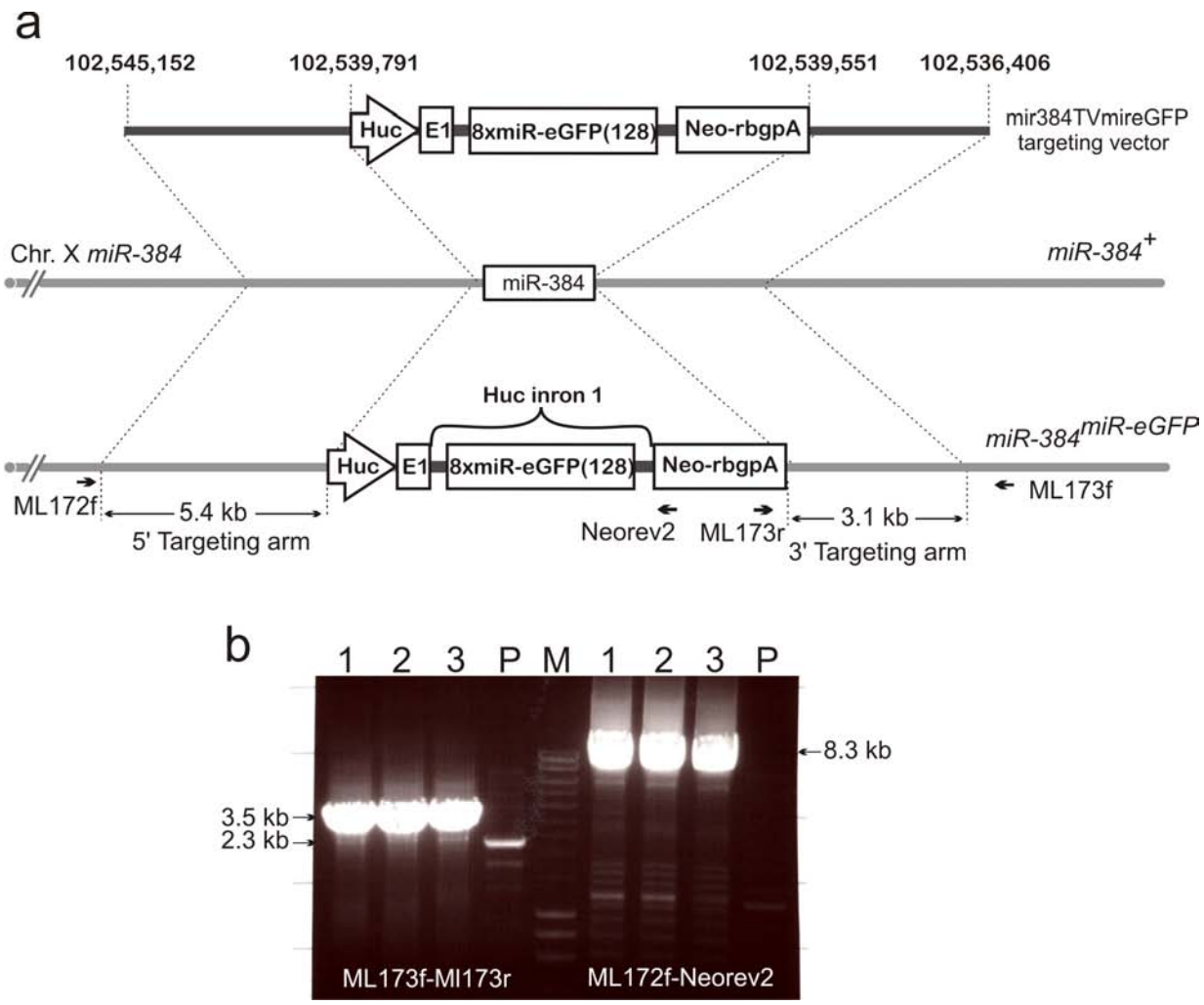
2.1.4. Generation of an artificial miR-eGFP knockin in the *Blm*$^{e/e}$*Hprt*^{PBin2} cell lines

In order to provide a uniform and constant level of miR-eGFP expression, a polycistronic expression cassette containing a Huc-driven miR-eGFP octamer with target sequence 128, residing within the intron of the Neomycin resistance gene was targeted to the *X*-linked *mir-384* locus. An *X*-linked locus was selected because this chromosome can not be lost in XY ES cells. The *mir-384* locus was selected because it is not expressed in mouse ES cells, although it is expressed in neurons (Marson et al., 2008). Therefore, disruption of *mir-384* is unlikely to result in an ES-cell phenotype.

A targeting vector miR-384TVmiR-eGFP was constructed by replacing the PGK-Puro cassette on the *mir-384* targeting vector with the Huc-driven miR-eGFP multimer and Neo resistance gene co-expressing cassette using recombineering (Details of the construction were described in methods in Chapter 2). The miR-384TVmiR-eGFP targeting vector was linearised with *Sbf*I and electroporated into 1x10^7 BlmHprtPBIn2 cells. The electroporated cells were selected with G418 for seven days and 48 colonies were picked and screened for correctly targeted events using long-range PCR (with primers ML173f and ML173r) across the short targeting

arm. This yielded three PCR-positive clones, giving rise to a targeting efficiency of 6 %. These three clones were further confirmed for the homologous recombination at the long targeting arm by long-range PCR (with primers ML172f and Neorev2). Figure 5-7 shows the gene targeting strategy and the long-range PCR confirmations of the correctly targeted clones.

Figure 5-7: Gene targeting of miR-eGFP multimer construct to the *miR-384* locus.



a, Gene targeting strategy with the targeting vector, wild-type miR-384 allele and the targeted allele shown. The genomic fragments used as targeting arms are indicated in the targeting vector with coordinates based on NCBI build37. b, Long-range PCR confirmation of correctly targeted clones for both targeting arms. P, the parental untargeted cell line. 1,2,3, three independent clones. M, Bio-Rad hyper ladder I marker.

2.1.5. Validation of the miR-eGFP knockin *Blm*$^{e/e}$*Hprt*^{PBin2} cell line

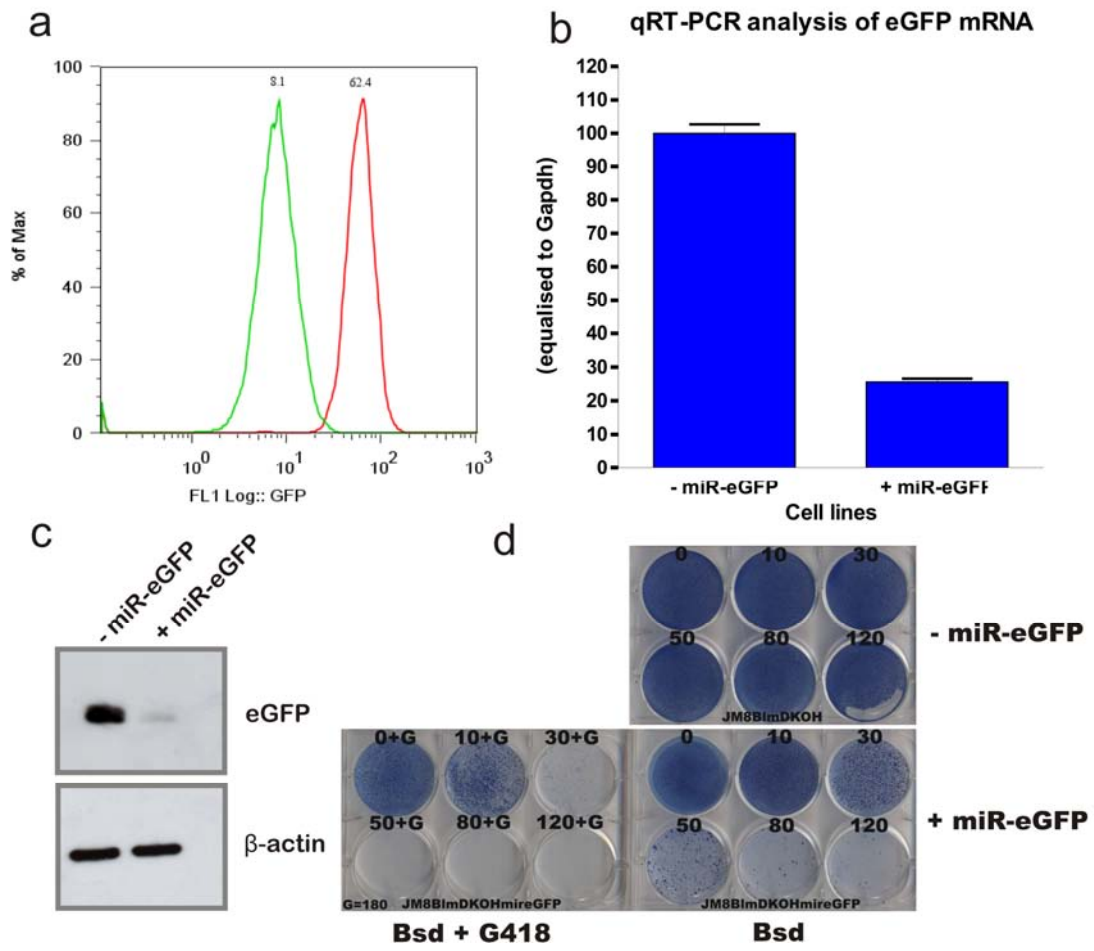
The three correctly targeted clones were subjected to karyotype analysis. Eight metaphases of each clone were examined. Two clones showed a normal karyotype with 40 chromsomes present in all metaphase spreads analysed and one clone had 40 % of the metaphases showing 39 chromosomes, possibly due to loss of the Y chromosome. One of the two clones with normal karyotype, B11, with the *Blm*$^{e/e}$; *Hprt*^{PBin2} background was chosen for subsequent experiments. The eGFP fluorescence of individual cells of this miR-eGFP targeted cell line was compared to the parental cell line without the miR-eGFP octamer. This analysis confirmed that the introduction of miR-eGFP resulted in approximately 90 % reduction in eGFP expression, Figure 5-8a. The eGFP mRNA and protein level was also confirmed to be significantly reduced, Figure 5-8b,c.

The repression of eGFP by a miR-eGFP with perfect complemtarity to eGFP should mediate mRNA cleavage and subsequent degradation. Since Bsd is translated from the same mRNA as eGFP, it was hypothesised that resistance to blasticidin would be reduced in miR-eGFP knockin cells. If this is the case, the screen can be conducted using blasticidin selection to eliminate the majority of the irrelevant cells in the mutant pools. A Blasticidin titration was conducted with concentrations ranging from 10 μg/ml to 120 μg/ml. The miR-eGFP knockin cells are more sensitive to high concentrations of blasticidin, with blasticidin concentration of 80 μg/ml and above clearing the majority of the cells. Although a few colonies arose from miR-eGFP knockin cells in high blasticidin concentrations, double selection with 180 μg/ml G418 and 80 μg/ml blasticidin completely cleared the wells. This observation suggests that these clones were derived from a very small proportion of cells in which the artificial miR-eGFP has been silenced. Therefore, dual blasticidin and G418 selection eliminate this false positive background.

Taken together, a reporter cell line has been successfully generated for probing the miRNA biogenesis pathway with a mRNA-mediated cleavage effector pathway. An artificial miR-eGFP was generated and the knockin of a single copy of this miR-eGFP octamer repressed up to 90 % of the eGFP expression in *Blm*$^{e/e}$ cells with two copies of an eGFP expression cassette. In

addition, miR-eGFP knockin cells are much more sensitive to high concentrations of the blasticidin than cells without miR-eGFP due to the miRNA mediated cleavage of the *eGFPIRESBsd* mRNA. This selection scheme provides a simple method to enrich homozygous mutants within mixed mutant pools.

Figure 5-8: miR-eGFP-knockin cell line validation.



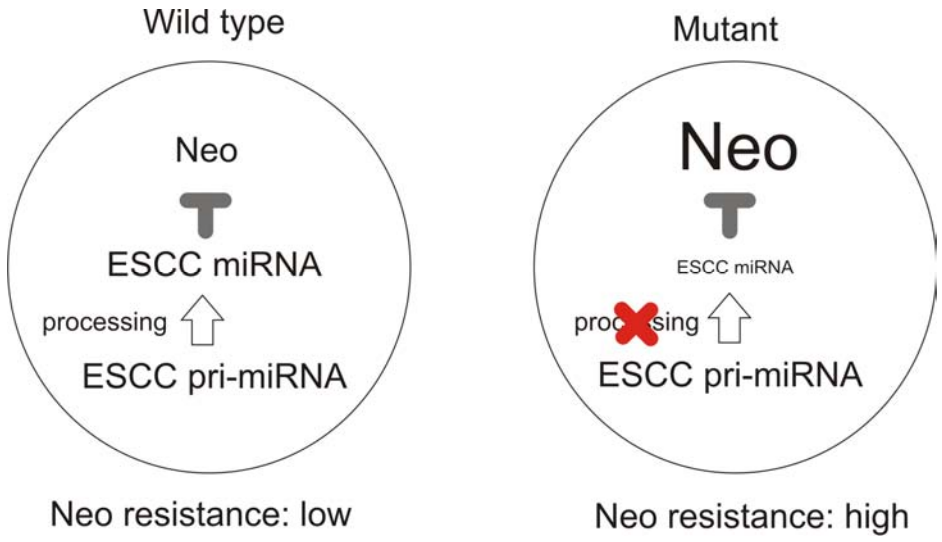
a, Fluorescence analysis of the eGFP intensity in miR-eGFP knockin cell line (green) compared to parental cell line (red). 10,000 events were gathered for the generation of the histogram. b, qRT-PCR (n=3) and c, protein level analysis of the miR-eGFP knockin cell line. d, Blasticidin resistant titration of the miR-eGFP knockin cell line (+miR-eGFP) compared to the parental cells without the miR-eGFP (-miR-eGFP). The blasticidin concentration used for each well was shown above with μg/ml as the unit. Blasticidin and G418 double-selection was also conducted with a constant 180 μg/ml G418.

2.2. Development of the endogenous miRNA target reporter

As previously described, a miR-eGFP system has been generated which probes one branch of the miRNA effector pathway. In order to assess the alternative effector pathway, translational repression, another reporter system is required. In lieu of an artificial miRNA reporter system, a highly expressed endogenous miRNA can be used and a reporter cassette can be generated to provide a selection scheme for mutant identification.

The mir-290 cluster is highly expressed in mouse ES cells comprising up to 70 % of all the miRNA expressed in ES cells (Marson et al., 2008). Four out of six mature miRNAs encoded within the mir-290 cluster, mir-291a-3p, mir-291b-3p, mir-294 and mir-295, share the same seed sequence AAGUGCU (Wang et al., 2008c). In addition, three miRNAs derived from the mir-320 cluster, mir-320b, mir-320c and mir320d, also share the same sequence, although the expression level of mir-320 is relatively low in mouse ES cells (Marson et al., 2008). These miRNAs are termed ES cell specific Cell Cycle regulating miRNAs (ESCC) miRNAs, and one of their functions is to promote the rapid G1-S cell-cycle transition in ES cells (Wang et al., 2008c). The abundance of the ESCC miRNAs with the identical seed sequence makes them an attractive candidate for use as a miRNA reporter, as the target suppression should be highly efficient due to this miRNA redundancy. Thus, a reporter cassette with the ESCC miRNA seed sequence recognition sites in the 3'UTR of the reporter can provide a direct repression of this reporter by the ESCC miRNA. In miRNA biogenesis mutants, the lack of miRNA elevates the reporter repression, providing a selection scheme for the screen. Figure 5-9 shows the reporter strategy.

Figure 5-9: An endogenous miRNA-based reporter system.

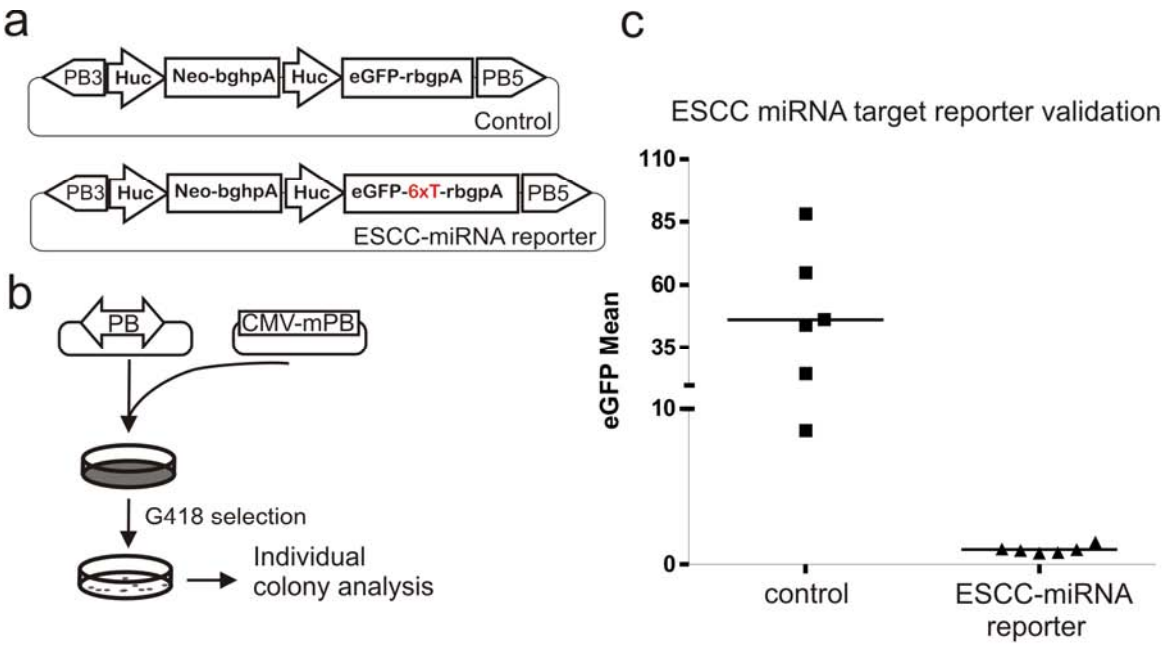


2.2.1. Artificial ESCC-miRNA target generation and validation

Several of the ESCC miRNA targets and target sites have been validated *in vitro*, including the cell cycle regulator, the inhibitor of the cyclinE/Cdk2 complex, Cdkn1a (p21) (Wang et al., 2008c). One of the ESCC-miRNA target sites within the 3'UTR of *Cdkn1a* and its 50 bp surrounding sequences were PCR-amplified and six tandem copies of the target sites were engineered in between the eGFP coding sequence and the polyadenylation signal to detect reporter knockdown by the endogenous ESCC-miRNAs, Figure 5-10a. The negative control in this experiment is an eGFP expression construct without ESCC target sites. The reporter expression cassette and the control were stably delivered into the ES cell genome by the PB transposon. Co-electroporation protocol used in this experiment was designed to deliver one copy of the transposon per genome and eGFP expression was examined in G418 resistant colonies, Figure 5-10b. AB2.2 cells ($1x10^7$) were co-electroporated with 100 ng PB transposon-containing plasmid and the mPBase-expression plasmid. The cells were selected with G418 for eight days and the colonies were picked, expanded and analysed using fluorescent activated cell sorting (FACs) for their eGFP intensity. Colonies derived from cells transfected with control PB plasmid gave rise to eGFP with varying intensities. This may be due to the locus-specific influence on eGFP expression. The colonies derived from cells transfected with the

ESCC-miRNA reporter PB plasmid expressed little eGFP. Therefore, the ESCC-miRNA reporter seems to be efficiently repressed by the endogenous ESCC miRNAs.

Figure 5-10: ESCC-miRNA eGFP reporter analysis.



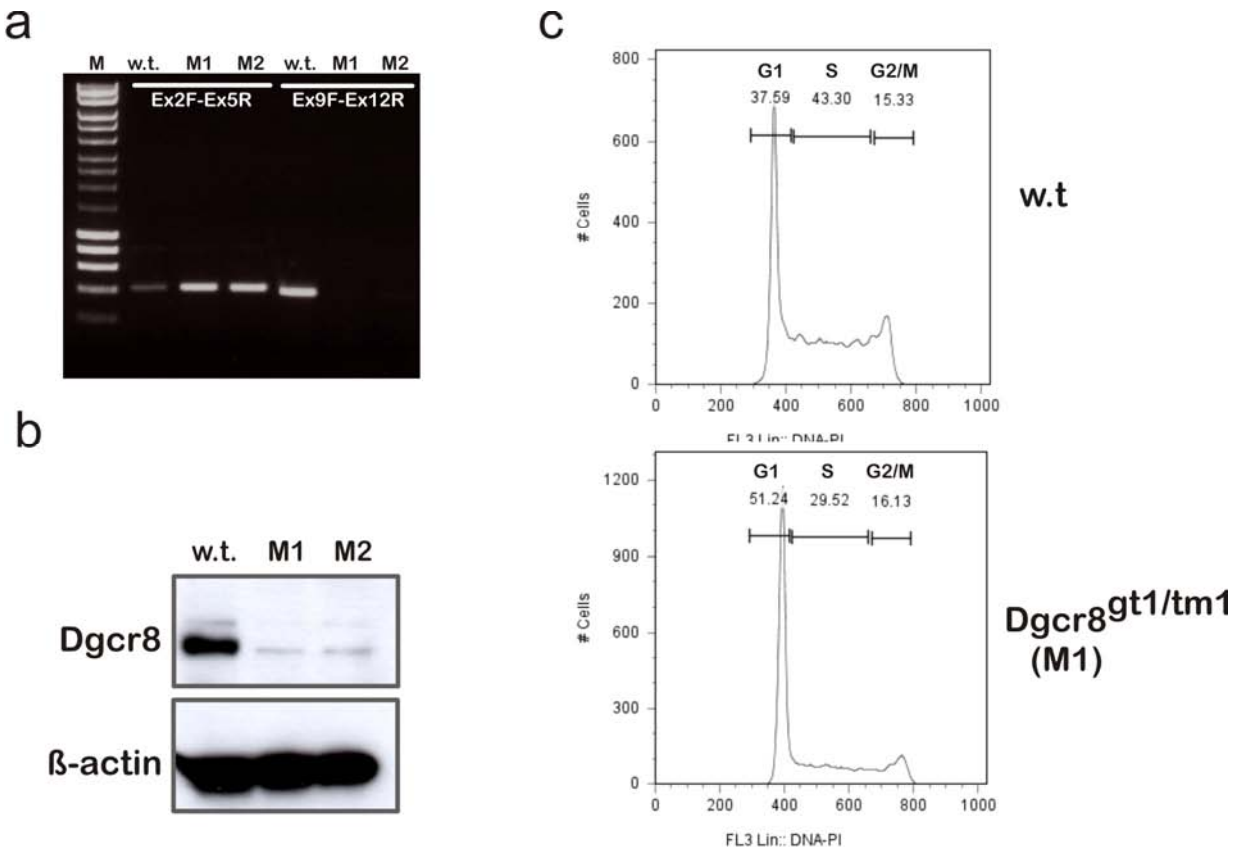
a, the ESCC-miRNA target reporter and the control in a PB transposon. b, PB transposon and transposase co-electroporation scheme to deliver the reporter and control constructs using single-copy PB delivery conditions. c, Analysis for eGFP intensity of six individual clones picked from each electroporation. The lines in the graph are the median values of the six clones.

In order to confirm that suppression of the ESCC-miRNA reporter is mediated by endogenous miRNA, the ESCC-miRNA reporter can be introduced into wild-type or *Dgcr8*-deficient ES cells. *Dgcr8*-deficient ES cells can not produce mature canonical miRNAs as *Dgcr8* is required for the Drosha processing to convert pri-miRNAs to pre-miRNAs (Wang et al., 2007). If the ESCC-miRNA reporter repression is mediated by the endogenous miRNA, the reporter repression should be abolished in *Dgcr8*-deficient cells, thus the reporter should be expressed, giving rise to eGFP fluorescence.

A *Dgcr8*-deficient ES cell line has been previously generated by Matthew Davis (Davis, 2009). One allele of the *Dgcr8* in a clone (XH157) isolated from BayGenomics resource with the *Dgcr8* allele disrupted by gene trapping with the *SA-βgeo-pA* trap cassette inserted in intron 8 of the *Dgcr8* (*Dgcr8*^{gt1} allele). The second allele was disrupted by insertional targeting of a SA-Hygromycin-pA cassette into intron 4 of the Dgcr8 locus and the resulting allele had a duplication for the genomic region from exon 2 to exon 6 (*Dgcr8*^{tm1} allele).

The Dgcr8 expression in this cell line was validated by RT-PCR and Western blotting to detect the production of the endogenous mRNA and protein, respectively. RT-PCR products amplified using primers Ex2F and Ex5R were detected in both wild-type and the double-allele trapped *Dgcr8* (*Dgcr8*$^{gt1/tm1}$) clones (M1 and M2) as both endogenous and chimeric trapped alleles produce mRNA including transcripts derived from these upstream exons. However, RT-PCR product amplified using primers Ex9F and Ex12R was not detected in mutant clones, as this portion of the transcript can only be detected in wild-type *Dgcr8* mRNA. Therefore, endogenous mRNA was not present in the *Dgcr8*$^{gt1/tm1}$ clones, Figure 5-11b. The endogenous protein level was assayed by Western Blotting and in *Dgcr8*$^{gt1/tm1}$ cells a faint Dgcr8 protein signal was detected, but the level of expression was significantly weaker than the wild-type cells, Figure 5-11c. Further analysis of the cell cycle was conducted to confirm the functional loss of the *Dgcr8*, as it was shown previously that an increased of G1 phase accumulation and a reduced proportion of S phase in *Dgcr8*$^{gt1/tm1}$ ES cells compared to wild-type cells. The *Dgcr8*$^{gt1/tm1}$ cell line showed a significant increase in the proportion of G1 phase population and reduced S phase population compared to the wild-type cells, Figure 5-11d. Illumina sequencing was conducted by Matthew Davis to examine all mature miRNAs in *Dgcr8*$^{gt1/tm1}$ cells compared to wild-type cells and *Dgcr8*$^{gt1/tm1}$ cells possessed a global reduction in canonical mature miRNAs (unpublished). Taken together, the *Dgcr8*$^{gt1/tm1}$ mutation is hypomorphic; however, this cell line behaves as a *Dgcr8* loss-of-function mutant.

Figure 5-11: *Dgcr8*$^{gt1/tm1}$ ES cell validation.

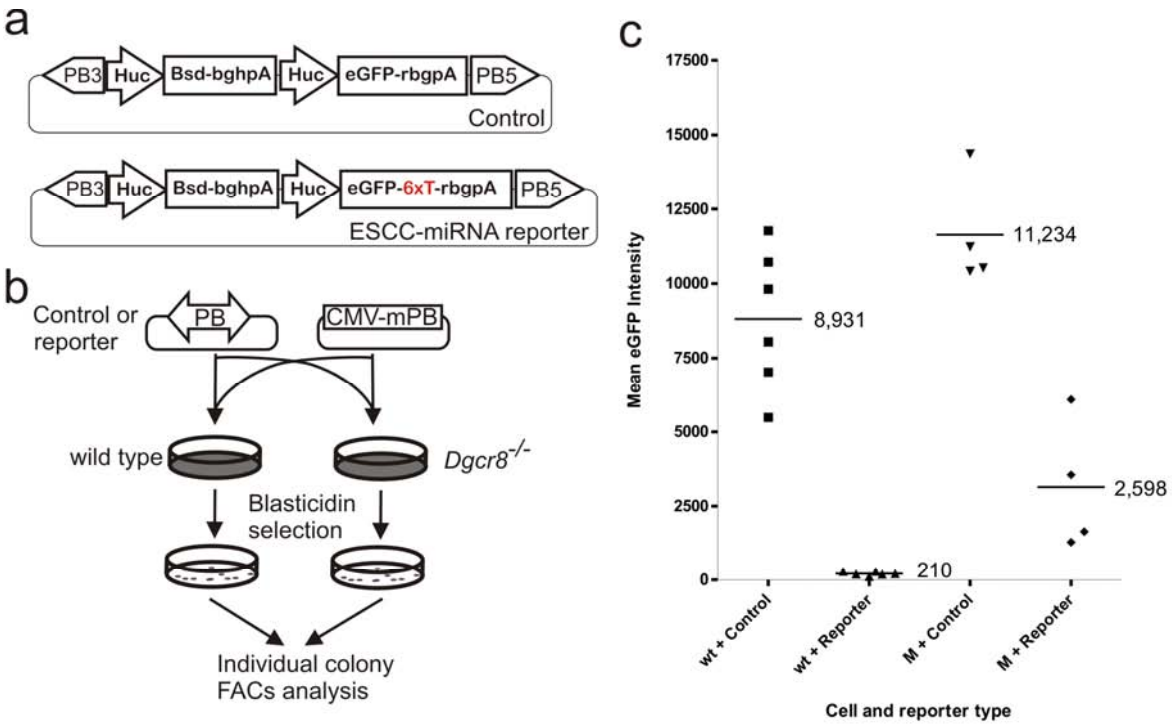


b, RT-PCR and c, western blotting to confirm the absence of endogenous *Dgcr8* mRNA and protein, respectively of the *Dgcr8*-deficient ES cell line (two clones M1 and M2). PCR with primers Ex2F and Ex5R can amplify products from both trapped transcript and the endogenous transcripts, whereas primers Ex9F and Ex12R can amplify only the wild-type transcript. c, cell-cycle analysis of the Dgcr8-null ES cells (bottom) compared to the wild-type cells (top). The percentage of cells in each phase of the cell cycle is shown in the histogram. The x-axis is the intensity of the propidium iodide (PI), which binds to the DNA and used to quantitatively measure the DNA content. The y-axis is the number of cells analysed and the total number of cells analysed is 10,000.

To confirm the causality of the endogenous miRNA mediated ESCC-miRNA reporter knockdown, the wild-type or *Dgcr8*$^{gt1/tm1}$ ES cells were co-electroporation of with either the control or ESCC-miRNA reporter-containing PB transposons. The eGFP intensity was compared in Bsd-resistant clones, Figure 5-12. As observed previously, in colonies derived from wild-type cells (n=6), the ESCC-miRNA reporter express little eGFP, whereas in colonies derived from *Dgcr8*$^{gt1/tm1}$ cells (n=4), an average over ten-fold eGFP expression was detected,

Figure 5-12c. The diverse levels of eGFP expression may be due to the locus specific influence on the eGFP reporter as the PB transposons integrate in different loci in all clones. However, the eGFP expression level of *Dgcr8*$^{gt1/tm1}$ cells transfected with the ESCC-miRNA reporter was four fold less compared to the control reporter without the ESCC-miRNA target sequence multimer in the 3'UTR of the eGFP coding region. Taken together, these results suggest that the ESCC-miRNA reporter was knocked down by the endogenous ESCC miRNAs in ES cells. However, the ESCC-miRNA reporter expression is reduced due to the presence of the repetitive ESCC-miRNA target sequences present in its 3'UTR. Despite this reduction in expression level of the reporter, the ESCC-miRNA reporter may still provide sufficient expression for screening purposes.

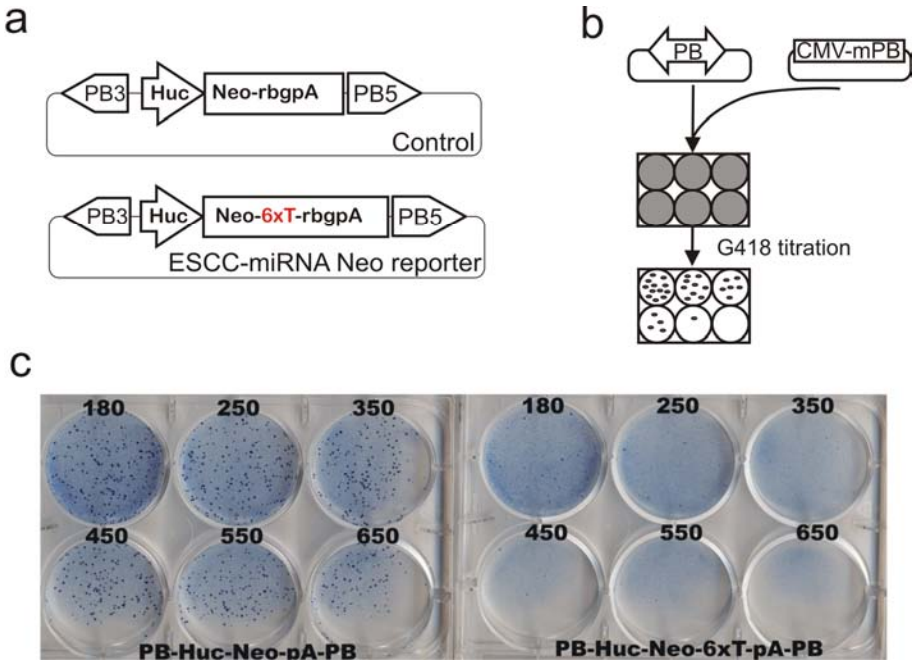
Figure 5-12: Causality establishment of the ES cell miRNAs and the ESCC-miRNA reporter.



a, the ESCC-miRNA Neo reporter and control constructs. "6xT" highlighted in red is the six copies of the ESCC-miRNA recognition sites. b, PB transposon and transposase co-transfection scheme to deliver the ESCC reporter and control constructs to both wild type or *Dgcr8*$^{gt1/tm1}$ cells. c, eGFP intensity measured in Bsd-resistant clones derived from the transfection. The median of each condition is shown on the diagram with a line presenting the median.

In order to test the feasibility of the ESCC-miRNA reporter can be differentially selected using a drug resistant cassette, a Neo reporter assay was also conducted to validate whether the ESCC-miRNA mediated reporter repression is selectable using G418. The assay was very similar to the eGFP-based system (Figure 5-10a) except that a Neo resistance gene was used in place of eGFP and the PGK-Neo cassette was deleted, Figure 5-13a. AB2.2 cells ($1x10^7$) were co-electroporated with 100 ng PB transposon-containing plasmids and the mPBase-expressing plasmid for the control and the reporter, independently. The electroporated cells were divided equally between six wells in a six-well culture plate. A G418 titration, ranging from 180 μg/ml to 650 μg/ml, was conducted on these cells, with each G418 concentration tested in one well of the 6-wells, Figure 5-13b,c. The cells transfected with control plasmid were resistant to G418 at all concentrations tested. However, the cells transfected with ESCC-miRNA Neo reporter were more sensitive to G418 than control even at the lowest concentration tested, Figure 5-13c. Thus, an ESCC-miRNA Neo-resistant reporter can be distinguished using G418 selection and this provide a screening scheme using the ESCC-miRNA reporter strategy. The previous assay using the *Dgcr8*$^{gt1/tm1}$ cells uncovered the reduced expression of the ESCC-miRNA reporter compared to the control reporter without the ESCC-miRNA target sequence in the 3'UTR, Figure 5-12. Therefore, further titration of the G418 concentration should be conducted in *Dgcr8*$^{gt1/tm1}$ cells to identify the level of G418 resistance of the reporter containing the ESCC-miRNA target sites. However, an unrecyclable Neo cassette is already present in the *Dgcr8*$^{gt1/tm1}$ cells, thus they are not suited for such an experiment. A *Dgcr8*-deficiency can be generated in the cell line containing the ESCC-miRNA reporter to further aid the titration of the G418 for the screening purposes.

Figure 5-13: ESCC-miRNA Neo reporter analysis.



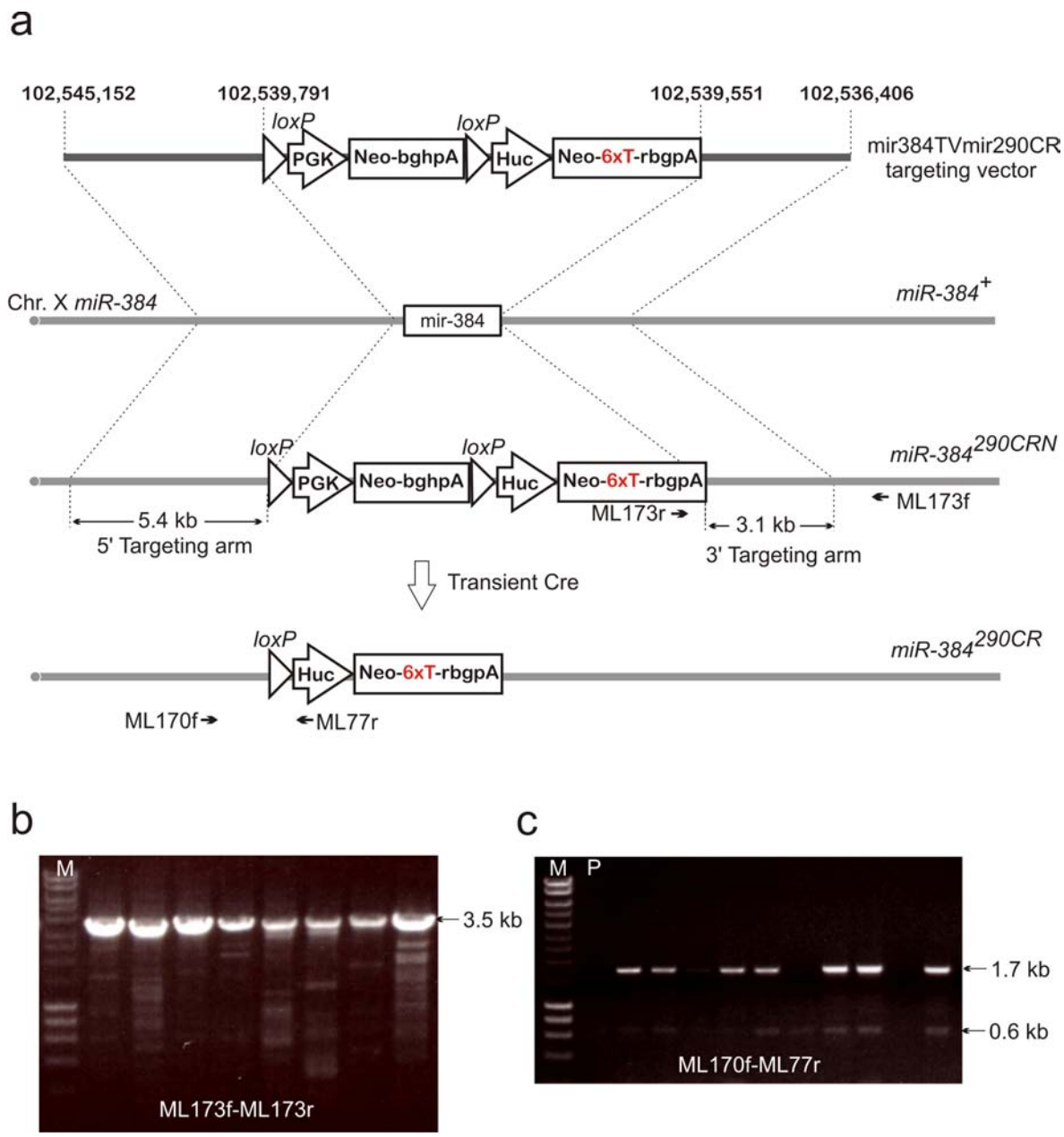
a, the ESCC-miRNA Neo reporter and control constructs. b, PB transposon and transposase co-electroporation scheme to deliver the ESCC reporter and control constructs using single-copy PB delivery conditions. Each electroporation was divided and plated equally in six wells of a 6-well plate and a G418 titration was conducted. c, The G418 titration. Left plate: control construct. Right: ESCC-miRNA Neo reporter constructs. The G418 concentration is indicated on top of each well (μg/ml).

2.2.2. Generation of an ESCC-miRNA target reporter knockin *Blm*$^{e/e}$*Hprt*^{PBin2} cell line

A stable reporter cell line was generated by targeting the ESCC-miRNA Neo reporter into the *X*-linked *mir-384* locus. The targeting vector, miR-384TVmir290CR, was constructed by introducing a *loxP*-flanked *PGK-Neo* cassette together with the ESCC-miRNA *Neo* reporter to the mir-384 targeting vector by recombineering (Chapter 2 methods). The *PGK-Neo* cassette provides a selection scheme for the gene targeting event and it can be removed subsequently by transient Cre expression, Figure 5-14a. The targeting vector was linearised with *Sbf*I and electroporated into 1x10^7 BlmHprtPBIn2 cells. The electroporated cells were selected with G418 and after seven days of selection, 48 colonies were picked and screened for correctly targeted events using long-range PCR (with primers ML173f and ML173r), Figure 5-14b. The targeting efficiency was 17 % in this experiment.

Eight metaphases of each correctly targeted clone were examined and four out of six clones had normal karyotype, with 40 chromosomes and the rest two clones had 39 chromosomes, probability due to the loss of Y chromosome. One karyotypically normal clone was chosen and $3x10^6$ cells were electroporated with pCAG-Cre plasmid to pop-out *PGK-Neo* cassette. Three days post-electroporation, the transfected cells were replated at 1,000 cells per 90 mm plate in duplicates. One of the 90 mm plate was selected under 180 µg/ml G418 and the other plate was cultured without drug selection. After eight days, colonies were picked from the plate without G418 selection. The pop-out positive clones (1.7 kb) were screened by genomic junction PCR using primers ML170f and ML77r. The cells without the *PGK-Neo* cassette popped out could not be amplified in this PCR condition, Figure 5-14c. The resulting reporter cell line will be tested to establish the minimal G418 concentration and can then be subsequently used for the screening.

Figure 5-14: The *miR-384* locus targeting with the ESCC-miRNA reporter construct.



a, Targeting vector, wild-type mir-384 allele, targeted allele (*miR-384290*^{CRN}) and the pop-out allele (*miR-384290*^{CR}) structures. The genomic fragments used as targeting arms are indicated in the targeting vector with coordinates based on NCBI build37. b, long-range PCR confirmation of correctly targeted clones using primers ML173f and ML173r. c, PCR screening of PGK-Neo cassette popout, and the 1.7 kb product represents the popout events. The 0.6 kb product is a non-specific PCR product. P, parental cell line, *miR-384290*^{CRN}, without *Cre* transfection. M, Bio-rad Hyper ladder I marker.

3. Discussion

This chapter describes the establishment and validation of two reporter systems which enable genetic screens for the miRNA biogenesis and its downstream effector pathways. These two reporter systems complement each other in the types of the mutants that can potentially be identified. Both systems cover the common upstream miRNA processing pathway, i.e. components that play a role in pri-miRNA to pre-miRNA processing and subsequent transport from the nucleus to the cytoplasm. On the other hand, they can also probe novel components that are specific for each of the two miRNA effector pathways, i.e. the mRNA cleavage- or translational repression-mediated gene repression.

It is thought that the degree of the complementarity between the mature miRNA sequence and its target mRNA distinguishes between these two branches of the effector pathways and different Ago proteins are central to the carry out the effector functions. As Ago2 is the only mammalian Ago protein out of the four homologues with the endonucleolytic ("slicing") function, perfectly complementing miRNA and its mRNA target can induce mRNA cleavage by the Ago2-containing RSIC complex. mRNAs recognised by imperfectly matched miRNAs, which are associated with Ago1, 3, and 4, can only mediate mRNA destabilisation and translational repression. However, Ago2 or Ago1, 3, and 4 are not exclusively associated with perfect and imperfect complementarity, respectively, between the miRNA and its target and it is not clear whether there is any novel component involved in guiding the choice. Furthermore, it is not yet known whether the two branches of the miRNA pathways recruit different components to conduct or regulate effector functions. In addition, downstream of RISC, little is known about how cells "treat" the miRNAs and whether the two types of effector pathway are differentially "treated". Finally, the biogenesis and regulation of the newly discovered endo-siRNAs is only just starting to be revealed in mammalian systems (Tam et al., 2008; Watanabe et al., 2008). The resemblance of post pri-miRNA processing of the artificial miR-eGFP system to endo-siRNAs may provide some insights into the biogenesis and downstream effector pathway of this new class of endogenous small RNAs.

The establishment of sensitive reporter systems is very helpful to "translate" non-visible phenotypes to selectable schemes, thus these systems expand the possibilities of investigating many biological pathways in cell culture. When used in conjunction with the *Blm*-deficient ES cells to conduct recessive genetic screens, such reporter systems are also useful in overcoming a major limitation in the mixed pooling strategy to isolate a few pathway-related homozygous mutants from a vast number of irrelevant cells mixed within the same pool. This situation is almost like finding needles (relevant homozygous mutants) in a haystack (irrelevant cells in the pool), but with a magnet (a sensitive reporter system), the needles can be easily found.

Targeting rather than random integration of the reporter was adopted to introduce reporters into the ES cell genome. There are two main reasons for doing so. Firstly, reporters randomly integrated in different genomic loci may show different levels of expression due to the influence of locus-specific chromatin structures and methylation status. Moreover, random integration of linearised transgenes is prone to silencing. Thus, a single clone with a good expression level has to be established in order to avoid variation in the expression level; as such variations can give rise to a significant background level during the screening procedures. Secondly, random integrations almost always land in autosomal loci, as 38 out of the 40 chromosomes are autosomes. In *Blm*-deficient ES cells, the LOH rate is much higher than wild type cells, with approximately one cell in 2,000 at a specific locus losing its heterozygosity every generation. Therefore, randomly integrated reporters can be lost during culture expansion in the *Blm*-deficient background, leading to false positive and negative clones. Gene targeting provides a control over the expression uniformity. In addition, targeting a reporter into both alleles of an autosomal locus, or by targeting the reporter to *X*-linked loci, avoids LOH-mediated reporter loss. Recombineering technology allows the rapid construction of targeting vectors from BACs or modification of existing targeting vectors. Thus, introduction of defined genetic modifications in mouse ES cells by gene targeting is very simple experimentally and provides significant advantages for reporter expression.

An artificial miR-eGFP has been successfully generated and the miR-eGFP multimers were placed within the intron of a neomycin expression cassette. The main advantage of this architecture is to maintain miR-eGFP expressing cells can be maintained using G418 selection. Silencing of the miR-eGFP could introduce false positive hits in the screen, and thus can be eliminated in this scheme. In addition, by using G418 in conjunction with the blasticidin selection, mutants that are defective in the miR-eGFP transcription can also be excluded.

The miRNA-reporter polycistronic expression strategy described here is very useful for other applications *in vitro* and *in vivo*. The artificial or even endogenous miRNAs can be inserted into any transcription unit with tissue specific or temporally controlled promoters. *In vivo*, the expression of such miRNAs can also be monitored by incorporating reporters which provide fluorescence or luminescence readouts. Since the roles of many miRNAs are unknown, this expression strategy can provide a system to investigate miRNA function by over-expression or ectopic-expression.

During the generation of the miR-eGFP system, several factors have been tested, including promoters with different strengths, miRNA backbones, and copy numbers of the miRNA structures, in order to produce the maximum degree of repression. It was found that the target sequence or its secondary structure is the major factor influencing the repression efficiency. In addition, different combinations of conditions tested gave rise to a wide range of repression but none were able to provide complete repression of eGFP. Such a range in target repression may have biological significance. miRNAs may repress their targets with a range of efficiencies depending on the target sequences *in vivo* and subsequently produce a range of protein levels of the targets. This provides an additional dimension in regulating the biological systems, as different concentrations of the protein products may give rise to different phenotypic outcomes. This kind of biological phenomenon has been most extensively demonstrated in pattern formation with a morphogen gradient during *Drosophila* embryogenesis for example. This strategy may be used by miRNA mediated gene expression control for tuning phenotypes. Finally, this incomplete knockdown effect further highlights

the fact that shRNA-mediated knockdown is equivalent to a “hypomorphic” mutant, thus the phenotypic interpretation of these experiments has to be cautious.

The difficulty of achieving efficient knockdown with an artificial miRNA stimulated me to try an alternative approach in which endogenous miRNAs with the identical seed sequence was utilised to repress a reporter with the target sites engineered into the 3’UTR. It was noted that the ESCC-miRNA reporter with multiple ESCC miRNA target sequence in the 3’UTR showed a reduced expression compare to the otherwise identical reporter without these elements. This could due to the decreased stability of the messenger RNA and consequently a reduced protein production. However, the ESCC-miRNA reporter expression level is ten fold higher in cells lacking of the ESCC miRNAs compared to wild-type cells. This different may be sufficient enough to distinguish homozygous miRNA-pathway mutants from the irrelevant cells mixed within the pool. A cell line has been constructed which contains the neomycin resistant gene with the ESCC miRNA target sites in the 3’UTR targeted to an *X*-linked locus in the BlmHprtPBIn2 cell line. Further G418 titration is required to identify the appropriate drug selection level for conducting the screen. The generation of a miRNA biogenesis mutant in this cell line can facilitate to achieve this goal.

A reporter with a specific miRNA target site present in the 3’UTR has been used previously as a miRNA decoy to dampen the effect of miRNAs on their endogenous targets, a technology termed “miRNA sponges”, although such methods can only achieve a partial elevation of the expression of endogenous miRNA targets (Ebert et al., 2007). A recent finding suggests that these types of miRNA “sponges” exist naturally (Poliseno et al., 2010). Transcribed pseudogenes can act as natural “sponges”. This finding provides evidence that pseudogenes play active roles in regulating gene expression, challenging the conventional thoughts that pseudogenes are genomic junk. Furthermore, this also highlights the fact that there may be many novel avenues towards regulating miRNA-mediated target repression.