Chapter 4

Identifying mediators of malignant transformation in cancer using genome-wide transposon-based gene activation

4.1 Introduction

The genome-wide CRISPR-Cas9 knockout screen described in chapter three has one clear limitation - it can only identify genes that cause transformation when subjected to loss-of-function mutations. Many existing mutations known to cause transformation affect oncogenes, that require overexpression, upregulation or activating point mutation to initiate tumourigenesis. These genes include those involved in pro-proliferative signalling such as *RAS* and *SRC* (Oneyama et al., 2007; Yamamoto et al., 1999), and genes where mutation allows the bypass of replicative senescence, such as *TERT*, the catalytic subunit of telomerase (Nault et al., 2014).

To detect genes where gain-of-function mutations are responsible for transformation, an approach was needed that could upregulate expression of genes genome-wide, and experimentally identify the genes of interest. Genome-wide transposon screening is a powerful approach that can be used to insert a desired sequence across the genome in a range of cell types. These insertion sites are easily recoverable due to specific sequences within the inserts, allowing the numbers of insertions at different sites in a cell population to be quantified by sequencing (Friedrich et al., 2017).

This approach has been used previously by Friedrich *et al.* (2017) for cancer gene discovery in mice, using *PiggyBac*-based transposons for genome-wide insertional mutagene-

sis. Insertions are quantified using QiSeq, which comprises DNA fragmentation by acoustic shearing, library preparation through modified splinkerette PCR (Devon et al., 1995), and custom Illumina sequencing. In this chapter, this method was used to recover insertions that transcriptionally upregulate expression of a downstream gene. The plasmid used was pPB-SB-CMV-puro-SD, which contains the human cytomegalovirus (CMV) promoter, flanked by *PiggyBac* and *Sleeping Beauty* sites (Tsutsui et al., 2015). The CMV promoter strongly upregulates downstream transcription leading to increased gene expression, aiming to model the gene amplifications seen in some tumours (Xia et al., 2006).

Similarly to the CRISPR-Cas9 screen discussed in chapter three, the principle of this screen was to induce genome-wide modifications of the NIH3T3 cells, then allow them to form transformed foci in culture. Cells containing alterations that induce transformation will therefore be overrepresented in the final population, and the numbers of mutations at different loci can be determined using sequencing to identify the genes responsible.

4.1.1 Aims

Overall aim: To identify putative oncogenes involved in malignant transformation in human cancer.

1. To identify genes that may mediate transformation *in vitro* using genome-wide transposon-based gene activation screening in NIH3T3.

2. To prioritise hits from this screen using mutation data from existing human cancer sequencing projects.

3. To functionally validate prioritised hits *in vitro.*

4.2 Materials and Methods

4.2.1 Materials

Plasmids

pCMV-hyPBase This plasmid was obtained from Dr. Kosuke Yusa at the Wellcome Sanger Institute (Yusa et al., 2011).

pPB-SB-CMV-puro-SD This plasmid was a gift from Professor Cyril Benes ((Tsutsui et al., 2015))

pBabe-puro Ras-V12 This plasmid was a gift from Professor Bob Weinberg (Addgene plasmid # 1768)

Cell lines

NIH3T3 NIH3T3 wild-type cells were obtained from the American Tissue Culture Collection (ATCC® CRL-1658™).

Reagents

Reagent	Manufacturer
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma Aldrich
Fetal bovine serum (FBS)	Gibco
Gentra Puregene kit	Qiagen
Lipofectamine 3000 kit (Lipofectamine 3000 reagent and P3000)	Thermofisher Scientific
Opti-MEM™ reduced serum media	Gibco
Penicillin, streptomycin and L-glutamine (100X, 50mg/mL)	Gibco
Puromycin	InvivoGen
Trypsin-EDTA (0.05%)	Gibco

Table 4.1: **Reagents used in the methods described in chapter 4**

4.2.2 Methods

Screen design

The aim was to acheive 100X coverage of the genome, with each gene being upregulated in a mean of 100 cells in the screen. To acheive this coverage, the rationale behind a previous screen using pPB-SB-CMV-puro-SD was used (Chen et al., 2013). Assuming that the number of insertions within a region is distributed according to the Poisson distriution, if the aim is for <5% of genes to have <100X coverage, the mean number of insertions upregulating expression of any given gene should equal 117 (see equation).

$$P(X \leq 99) = 0.05$$

$$\sum_{x=0}^{x=99} e^{-\lambda} \frac{\lambda^x}{x!} = 0.05$$

$$\lambda = 117$$

Following the assumption of Burgess et al. that the CMV promoter can upregulate transcription of a gene when <64kb upstream of the transcriptional start site, 117 insertions within this region equates to a mean 0.547kb gap between insertions. Therefore, in the 3 million kb mouse genome, 5,484,375 insertions were required. These are only functional if on the coding strand, therefore 10,968,750 insertions were needed in total.

A 1:10 ratio of *PiggyBac* transposase plasmid to *PiggyBac* transposon plasmid has been determined to generate an expected 1-10 insertions per cell (Wang et al., 2008), providing a reasonable number of insertions without causing excessive cell lethality. Making the conservative assumption that one insertion per cell is generated, 10,968,750 transfected cells would represent the same number of insertions. Given a measured transfection efficiency of NIH3T3 of 23.5% (appendix B.7), this equates to 46,675,532 cells. To account for cell loss during processing, this was rounded to 5 x 10^7.

Transfection

Day 0 5 x 10^7 NIH3T3 wild-type cells were seeded in 26 15cm-diameter culture dishes at a density of 96,100 cells/mL in complete DMEM (DMEM supplemented with 10% FBS and 500μg/mL penicillin, streptomycin and L-glutamine). For the positive and negative controls, a 6-well tissue culture plate was seeded at a density of 50,000 cells/mL, in 2mL of complete DMEM per well.

Day 1 Media was changed to Opti-MEM™ reduced serum media before transfection. Cells were transfected using Lipofectamine 3000 according to the manufacturer's instructions, using the following reagent quantities (table 4.2). Cells in the 15cm dishes were transfected with pPB-SB-CMV-puro-SD and pCMV-hypBase, and for the positive control, 3 wells of the 6-well plate were transfected with pBabe-puro Ras V12. For the negative control, 3 wells of the 6-well plate were mock transfected, with the plasmid replaced with an equivalent volume of Opti-MEM™.

Reagent	Quantity per well (1 x 10^6 cells)	Quantity per dish (1.92 x 10^7 cells)
Lipofectamine 3000 Reagent	1.5µL	28.8µL
P3000	1µL	19.2µL
pPB-SB-CMV-puro-SD	0.45µg	8.64µg
pCMV-hypBase	45ng	0.864µg
pBabe-puro Ras V12	0.5µg	9.2µg

Table 4.2: **Transfection reagent quantities used in genome-wide transposon-based gene activation screen**

Day 2 16 hours post-transfection, media was changed to 20mL complete DMEM per dish, or 2mL per well.

Day 4 Media was changed to complete DMEM containing puromycin (2µg/ml). Media was changed every 3-4 days.

Day 30 Cells from the dishes were harvested by scraping, centrifuged (200*xg,* 5 minutes), washed with PBS and frozen at -80°C. The control cells were fixed for 1 hour using methanol, stained with 1% aqueous crystal violet for 10 seconds, washed with MilliQ water and air-dried.

DNA extraction

Genomic DNA was extracted using Qiagen Gentra Puregene kit according to the manufacturer's instructions.

QiSeq

Transposon insertions were quantified using QiSeq (Friedrich et al., 2017). Library preparation was performed with the assistance of Dr. Jonathan Cooper from the haematological cancer genetics group at the Wellcome Sanger Institute.

4.3 Results

At 30 days there were no or few foci in the mock-transfected wells, confirming the low rate of background transformation in NIH3T3. The *RAS*-transfected positive control contained a variable number of foci per well, ranging from three to approximately 45 in total. However, the overall increased focus formation compared to the control indicates that these cells can

be transformed by known oncogenes (figure 4.1). Many foci were seen in the transposon-transfected dishes, suggesting successful transformation by transposon-mediated gene over-expression.

The results of sequencing the insertions to determine their locations have not been returned yet. Once the sequencing data is generated the aim is to analyse it using the method detailed by Friedrich *et al*. to extract the insertion sites with the most reads and correlate these with the downstream genes they are likely to induce overexpression of. Putative tranformation-associated genes will then be prioritised using existing cancer genome data and validated using an arrayed focus formation assay as decribed in sections 3.2.2 and 3.2.2 of chapter three.

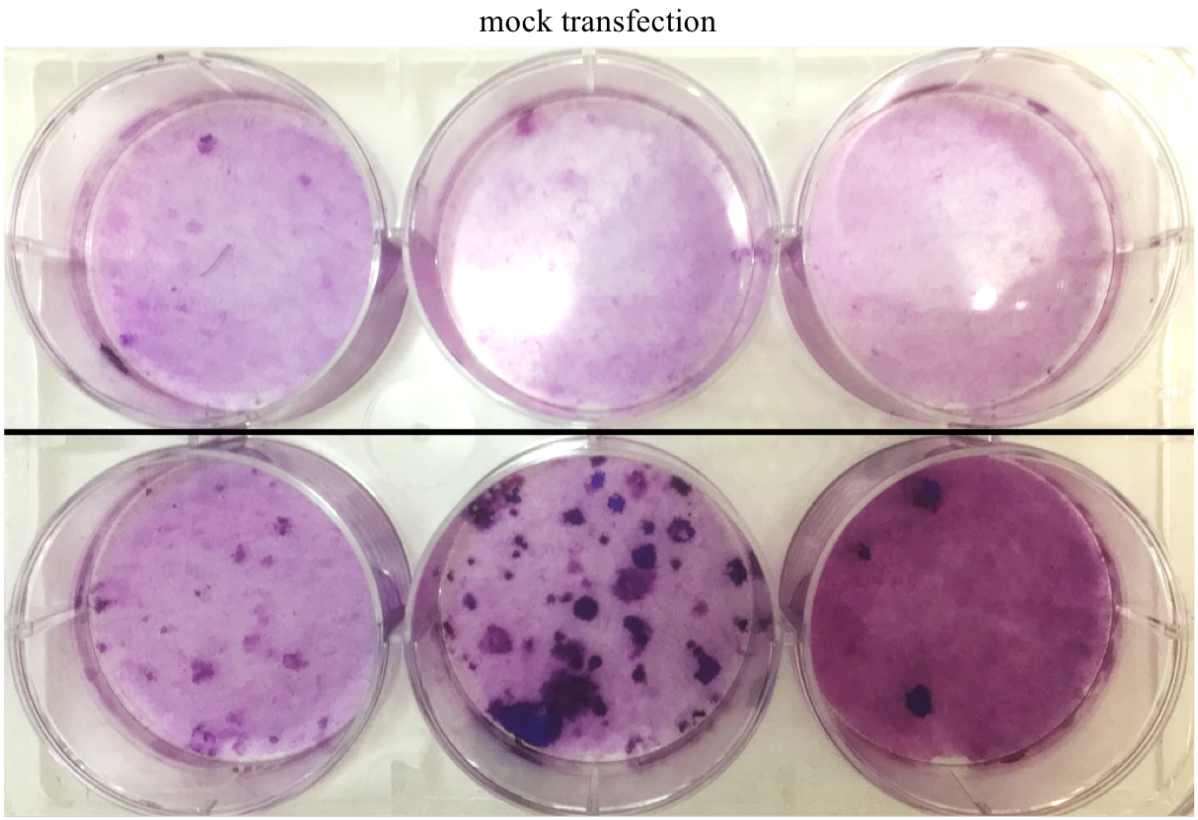

Figure 4.1: **Focus formation assay for negative and positive controls of the genome-wide transposon-based activation screen**

NIH3T3 cells were mock transfected as a negative control and transfected with a plasmid carrying known oncogene *RAS* (pBabe-puro Ras-V12) as a positive control for the genome-wide transposon-based activation screen. After culturing for 30 days, cells were stained using 1% aqueous crystal violet (see section 4.2.2).

4.4 Discussion

One limitation of this approach is that oncogenic point mutations are not accurately modelled, with the transposon insertions more closely recapitulating oncogene activation by amplification or upregulation. However, this may not preclude the identification of genes usually activated by point mutation, as many oncogenes can be be activated in multiple ways, giving similar phenotypes. For example, *RAS* is most frequently point mutated in tumours, but can also cause transformation by overexpression *in vitro* or amplification *in vivo* (Pierceall et al., 1991; Pulciani et al., 1985). There may also be issues inherent to the use of a cell-based model, as differences have been observed between the level of gene expression required for transformation *in vitro* when compared to that seen *in vivo*. For example, it has been shown that mutant human *RAS* requires over 100-fold higher expression to cause tranformation in cell lines compared to that seen in human cancers (Hua et al., 1997). However, *RAS* overexpression was used as the positive control in this screen and successfully induced transformation, which is promising.

As discussed in chapter three, the heterogeneous genetic background and complex karyotype of NIH3T3 may affect the nature of the genes identified from this screen, as the ability of a mutation to induce tranformation is likely to be dependent on other co-occurring mutations. Another limitation of the model discussed in chapter three is the absence of factors that play a part in early tumourigenesis *in vivo,* such as the immune microenvironment. As with the CRISPR-Cas9 screen, some of these issues could be resolved by future validation of any hits using mouse models. This could be done by injection of CRISPR-edited cells, or alternatively using a transgenic model that can activate the desired putative oncogene in a lineage- and time-specific manner (Blanpain, 2013).