Chapter 5

Conclusions and further directions

In this thesis I have described the genetic characterisation of the cell-based model of malignant transformation NIH3T3, and karyotyping of its daughter cell line NIH3T3-Cas9 followed by presenting two complementary forward genetic screening approaches used to identify putative oncogenes and tumour suppressor genes involved in the earliest stages of tumourigenesis.

5.1 Characterisation of the genetic background of NIH3T3 and NIH3T3-Cas9

The characterisation of the NIH3T3 genetic background consisted of whole-genome sequencing to identify SNVs and indels, comparison of these variants with known cancer-associated genes from the CGC and mutations from COSMIC, and characterisation of large-scale genomic alterations using M-FISH. 88 SNVs and indels with coding effects in CGC genes were identified; however determining the phenotypic effects of these variants and how they may influence transformation sensitivity was challenging. Some of the indels identified were predicted to have loss-of-function effects, but for many it was unclear whether this would promote tumourigenesis given the function of the gene and the mutations reported in human cancers. However the indel affecting a splice acceptor site in *Il6st* is worth further investigation due to its potential effect on known cancer-associated pathway JAK-STAT3 (Avalle et al., 2017; Hibi et al., 1990). The SNVs identified were not located at mutation hot-spots catalogued by COSMIC, however it is difficult to predict the effects of point mutations on a protein from sequence data alone. It may be possible to investigate this using computational approaches to the prediction of the effects of codon changes on protein structure and function (Tang and Thomas, 2016).

Additionally, mutations in genes that were not investigated in the analyses carried out in

this project may contribute to the transformation sensitivity of this cell line. For example, there may be murine genes without human homologues that affect transformation, or genes with unknown or poorly characterised effects on tumourigenesis. To ensure that these genes are also investigated, additional literature sources could be included, such as the positively selected driver mutations reported by Martincorena *et al.* in 2017. The transformation-sensitive phenotype of NIH3T3 could also be explained by epigenetic effects. To investigate this possibility, the epigenome of the cell line could be characterised using methods such as cytosine methylation profiling, and compared to epigenome data from patient tumour samples or cancer cell lines (Ehrich et al., 2008).

The analysis of large-scale genetic alterations in the daughter cell line NIH3T3-Cas9 showed an abnormal, predominantly tetraploid karyotype with whole chromosome amplifications, deletions and translocations. The effects of these alterations on cancer-associated genes may explain the transformation-sensitivity of the line. These results also revealed high levels of inter-cell heterogeneity, indicating genetic instability and suggesting that the cell line is continuing to evolve at the chromosomal level. This is also supported by the differences between the karyotype determined in this analysis and in previous cytogenetic characterisation by Leibiger *et al.* (2013). This finding has important implications for the use of this cell line as a model. As an enabling characteristic of the 'hallmarks of cancer' (Hanahan and Weinberg, 2011), genetic instability may influence transformation-related phenotypes and may may contribute to the transformation-sensitivity of these cells. Genetic heterogeneity may also affect the outcome of genetic screens, as experimentally-induced mutations do not act in a consistent genetic background given the polyclonality of the model.

In future, it would be informative to characterise the karyotype and genetic heterogeneity of the parental cell line NIH3T3 wild-type. These results could be compared with NIH3T3 samples from other sources to get a more complete picture of the heterogeneity present in the cell line as a whole. Large-scale alterations found consistently in the cell line could then be investigated at higher resolution to identify if any known oncogenes or tumour suppressor genes are amplified or deleted.

Overall, this work has made some progress in identifying possible reasons behind the transformation-sensitive phenotype of NIH3T3, while providing further evidence that cell lines evolve, and should not be assumed to remain genetically or phenotypically identical over time in culture.

5.2 Identification of candidate genes associated with malignant transformation

The genome-wide CRISPR-Cas9 knockout screen sought to identify novel candidate tumour suppressor genes that induce the formation of tranformed foci of proliferation when knocked out in NIH3T3. This screen successfully identified the known cancer-associated genes *Gnas, Kdsr, Sufu, Nf2, Cltc, Ptch1, Nf1* and *Pten*, along with putative novel candidates. The combined results of manual investigation of the normalised read counts and their subsequent analysis using MAGeCK showed that using the original gRNA sequence library as a control was the most successful approach for identifying gRNA seqeunces that were genuinely enriched, indicating that the corresponding genes may be involved in the formation of the transformed foci. This also highlighted an issue with using MAGeCK analysis to compare two samples that have been grown in culture, which can lead to the mis-identification of essential genes as enriched hits due to high statistical noise at low read counts. In future this could be avoided by removing gRNAs with control read counts below a certain threshold from the analysis, and by manually investigating the normalised read count data at an earlier stage to discard spurious hits. Alternatively, the screen could be redesigned to use only the gRNA sequence library as control.

The novel candidates identified from the comparison between the focus formation sample with the gRNA sequence library were *Rnf146, Kirrel, Mrgbp, Pdcd10, Traf3* and *Ube2m*. *Rnf146, Kirrel* and *Mrgbp* are described as putative oncogenes or reported as overexpressed in cancer samples (Gao et al., 2014; Yamaguchi et al., 2011; Zhang et al., 2018), whereas *Pdcd10, Traf3* and *Ube2m* are described as having tumourigenic effects when inactivated (Cukras et al., 2014; Hajek et al., 2017; Lambertz et al., 2015). This is interesting considering that all of these genes emerged from a screen employing loss-of-function mutation, suggesting that the effects of *Rnf146, Kirrel* and *Mrgbp* may be poorly characterised or context-dependent.

The planned analysis of the data generated from the transposon-based activation screen should generate candidate oncogenes involved in transformation. Once these hits and those from the CRISPR-Cas9 screen have been validated *in vitro,* successful hits could be verified *in vivo* using mouse models. A limitation of the current approach is that murine cancer-related pathways may show important differences from the human equivalents. To test genes in a human context, it would be informative to repeat any successful validation experiments using the human homologues of candidates in the human non-tumorigenic immortalised cell line MCF-10A (Qu et al., 2015). Following validation, further work to dissect the pathways containing these genes and their biological functions would be enlightening. Existing data on human cancer genomes from sources such as COSMIC could be used to investigate in which

tumour types these genes are commonly mutated, what kinds of mutations are most common, and which other mutations are significantly co-occuring or mutually exclusive. Overall, the further investigation of these candidates has the potential to inform the basic biology of malignant transformation and possible future therapeutic targets.